

Alternative Receptive Language Assessment Modalities and Stimuli for Children with ASD who  
are Minimally Verbal

By

Kristen Muller

Submitted to the graduate degree program in Speech-Language-Hearing: Sciences and Disorders  
and the Graduate Faculty of the University of Kansas in partial fulfillment of the  
requirements for the degree of Doctor of Philosophy.

---

Chair: Dr. Nancy Brady

---

Dr. Debora Daniels

---

Dr. Jane Wegner

---

Dr. Claudia Dozier

---

Dr. Kandace Fleming

Date Defended: June 5, 2019

The dissertation committee for Kristen Muller certifies that this is the approved version of the  
following dissertation:

Alternative Receptive Language Assessment Modalities and Stimuli for  
Children with ASD who are Minimally Verbal

---

Chair: Dr. Nancy Brady

Approved: June 5, 2019

## **Abstract**

There are very few receptive language measures that are appropriate for children with autism spectrum disorders (ASD) who have minimal verbal skills. The primary aim of this study was to evaluate and compare alternative modalities and stimuli used to measure receptive language skills in children with ASD who are minimally verbal. This study systematically compared participants' outcomes on three different receptive language assessment conditions: a picture-based standardized assessment, a touchscreen assessment, and an assessment that used real-object stimuli. A secondary aim of this study was to examine how individual characteristics such as cognition, maladaptive behavior, autism symptomology, and gesture use impact performance on assessment conditions. A tertiary aim involved the use of an intermodal preferential looking paradigm (IPLP) to determine whether eye gaze data may provide additional, implicit information about word comprehension for children with ASD. Twenty-seven students between the ages of 3 and 12 who had minimal verbal skills and a diagnosis of ASD participated in this study. Results from a crossed-random effects model showed that participants' scores in the real-object assessment condition were significantly higher than in the standardized condition, and marginally higher than scores in the touchscreen assessment. Together, cognition and fine motor skills accounted for 44% of the variance in participants' scores. IPLP data revealed that participants spent more time looking at target stimuli rather than foil stimuli in 36% of no-response trials. Clinical implications of these findings are discussed.

## TABLE OF CONTENTS

INTRODUCTION.....	1
Issues with Current Receptive Language Assessments.....	2
Direct Assessments.....	3
Indirect Assessments.....	6
Alternative Receptive Language Modalities and Procedures.....	8
Touchscreen Assessments.....	9
Real-Object Stimuli.....	10
Alternative Receptive Language Assessments.....	10
Eye-Gaze.....	11
Current Study.....	14
METHOD.....	16
Participants.....	16
Diagnostic, Cognitive, and Behavioral Measures.....	17
Childhood Autism Rating Scale-Second Edition.....	17
Mullen Scales of Early Learning.....	18
Caregiver Questionnaire.....	18

Experimental Assessment Conditions.....	20
Standardized Assessment Condition: PPVT-4.....	20
Touchscreen Assessment Condition.....	21
Real-Object Assessment Condition.....	22
Procedures.....	23
Experimental Assessments.....	24
Data Collection.....	25
Experimental Assessments.....	25
Eye-Tracking.....	26
Problem Behavior & Assessment Duration.....	27
Inter-Rater Reliability.....	27
Fidelity.....	29
Data Analysis.....	29
RESULTS.....	31
Performance on Assessment Conditions.....	31
Clinical Factors.....	34
Individual Characteristics.....	36

Eye-Gaze.....	38
DISCUSSION.....	40
Performance on Assessment Conditions.....	40
Clinical Feasibility.....	43
Individual Characteristics.....	45
Eye-Gaze.....	46
Clinical Implications.....	48
Assessment.....	48
Intervention.....	49
Attention.....	49
Limitations.....	50
Future Research.....	52
REFERENCES.....	55
APPENDIX A.....	64
APPENDIX B.....	66
APPENDIX C.....	67
APPENDIX D.....	68

APPENDIX E.....	69
-----------------	----

According to the Centers for Disease Control and Prevention (CDC), one in every 59 children has a diagnosis of autism spectrum disorder (ASD) (Baio et al., 2018). Many children with ASD have language delays, and about 30% remain completely nonverbal or minimally verbal (Plesa Skwerer, Jordan, Brukilacchio, & Tager-Flusberg, 2016). The term "minimally verbal" describes individuals who are school-aged or older who do not use phrase-speech spontaneously or functionally on a daily basis or who have approximately 30 or fewer spoken words that are used communicatively (Kasari, Brady, Lord, & Tager-Flusberg, 2013; Plesa Skwerer et al., 2016). Language assessments are a vital component in the diagnosis and treatment of ASD because language impairments are often a key feature of an autism diagnosis; and language status is a strong predictor of future language skills in children with ASD (Barbaro & Dissanayake, 2012; Muller & Brady, 2016; Volden et al., 2011). In addition, speech-language pathologists (SLPs) often use language assessments to inform therapy goals, and to track progress on these goals. Therefore, it is crucial for researchers and practitioners to have valid and accurate assessments that measure early language skills.

Unfortunately, language abilities in individuals with autism can be difficult to assess, both in direct standardized assessments and in indirect parent and caregiver questionnaires (Charman, 2004; Charman et al., 2003; Plesa Skwerer et al., 2016). Language comprehension, or receptive language skills, are particularly difficult to assess in this population due to behavioral and motivational issues (Charman, 2004; Plesa Skwerer et al., 2016), difficult response demands, such as requiring a child to point (Charman et al., 2003), and validity issues with parent reports (Bruckner et al., 2007; Tomasello & Mervis, 1994). Another primary challenge is that individuals with ASD are not included in the norms of most standardized language assessments



(Kasari et al., 2013). Therefore, results from these measures may not be valid for individuals with ASD.

Many studies suggest that children with ASD have greater impairments in receptive language than expressive language (Barbaro & Dissanayake, 2012; Ellis Weismer, Lord, & Esler, 2010; Hudry et al., 2010). Thus, receptive language assessment and intervention should be a primary research target for this population. Unfortunately, due to the paucity of appropriate assessment measures and behavioral challenges that often occur during assessment, individuals with ASD who are minimally verbal are often excluded from research (Kasari et al., 2013; Plesa Skwerer et al., 2016). However, recent literature suggests that alternative assessment modalities and procedures may yield more accurate assessment results for this population than traditional receptive vocabulary assessments (Plesa Skwerer et al., 2016). The focus of the current study was to evaluate and compare alternative assessment modalities and procedures used to measure receptive vocabulary skills in children who are minimally verbal.

### **Issues with Current Receptive Language Assessments**

Kasari and colleagues (2013) examined the psychometric properties and age ranges of current language assessments to determine their appropriateness for individuals who are minimally verbal and have ASD. Unfortunately, only one receptive language measure, the Peabody Picture Vocabulary Test Fourth Edition (PPVT - 4; Dunn & Dunn, 1997), was deemed well suited for use with this population (Kasari et al., 2013). Practitioners are encouraged to use multiple assessment measures in order to gain an accurate and thorough picture of a child's language skills (Camilleri & Botting, 2013; Conti-ramsdén & Durkin, 2012; Hudry et al., 2010; Kasari et al., 2013); however this is often challenging due to the multitude of issues in both

direct and indirect assessments when used with children with ASD who have minimal verbal skills.

**Direct Assessments.** Direct assessments use standardized procedures where every individual is assessed in a similar manner. This allows for more accurate comparison of scores across examinees. However, using direct measures with individuals with ASD is often difficult due to floor effects (Plesa Skwerer et al., 2016; Volden et al., 2011) and behavioral issues such as noncompliance, lack of motivation, and short attention span (Prelock & McCauley, 2012, Chapter 6). Also, individuals with ASD may not perform well under standardized testing conditions, because these novel situations may induce anxiety (Kasari et al., 2013).

**Floor effects.** Floor effects are seen when individuals do not receive many, or any, points on an assessment; and in turn, earn a score in the very lowest range of possible values for that assessment. These scores often pose challenges for both researchers and practitioners. Floor effects make it difficult to compare outcomes between different standardized measures, and to determine a profile of skills that are relative areas of strength and weakness.

Unfortunately, individuals with ASD often show floor effects on standardized language assessments. Volden and colleagues (2011) found that in their sample of 249 children with ASD, almost 30% scored at the floor of the Preschool Language Scale (PLS-4; Zimmerman, Steiner, & Pond, 2002). Similar issues were found with the Reynell Developmental Language Scales (RDLS; Edwards et al., 1997). The RDLS examines language skills beginning at 18 months; yet Charman and colleagues (2003) found that 22% of their sample of children with ASD, at 42-months old, scored below the basal score on the RDLS-III. Additionally, Kjelgaard and Tager-Flusberg (2001) found that 51% of their 89 participants, who had ASD and were between 4- and

14-years old, were unable to complete the Clinical Evaluation of Language Fundamentals (CELF; Wiig, Secord, & Semel, 1992).

As mentioned earlier, the PPVT is the only receptive language assessment recommended for use with individuals with ASD who are minimally verbal; however, researchers have found that this population is often unable to complete the training items on this assessment or score at the floor of the assessment. When individuals are unable to complete training items, the assessment is discontinued and no score is derived. In a recent study, Plesa Skwerer (2016) found that out of 18 children and teens who were minimally verbal with ASD, 6 were unable to complete the PPVT training items (32%), and 5 achieved a standard score on the floor of this measure (28%). Similarly, Brady, Thiemann-Bourque, Fleming, and Matthews (2013) found that nearly half of their participants who were 4- to 7-years old who had ASD and minimal verbal skills were unable to complete the training items on the PPVT. Also, in our ongoing research, we found that 54% of participants who were minimally verbal and between the ages of 3 and 60 were unable to pass the training items on the PPVT, and thus were unable to obtain a score (Muller & Brady, 2016).

Individuals with ASD may also be more likely to show floor effects on receptive language subtests than on expressive language subtests. Hudry and colleagues (2010) found that, in their sample of 152 toddlers with ASD, 43% scored on the floor of the receptive language subtest of the PLS-4 while only 32% scored at the floor of the expressive language subtest. Charman and colleagues (2003) also found that individuals with ASD are more likely to show floor effects on comprehension scales than expressive scales of the RDLS. These results signify the necessity for new measures and techniques that examine receptive language skills for individuals with minimal verbal skills and individuals with ASD. However, it is unclear whether

floor effects are a true representation of severe language deficits, or if current language assessments include tasks that are particularly challenging for individuals with ASD.

***ASD-related performance difficulties.*** Many receptive language assessments, such as the PPVT-4, the CELF-5 (Wiig, Semel, & Secord, 2013) and the PLS-5 (Zimmerman, Steiner, & Pond, 2011), require examinees to point to a referent after it is named (e.g., Point to dog; Point to I can eat this). These tasks may be difficult for individuals with noncompliance issues, motor delays, or limited behavioral repertoires (Brady, Anderson, Hahn, Obermeier, & Kapa, 2014). As mentioned earlier, individuals with ASD develop gestures, such as pointing, later than typically developing peers; and they use these gestures less frequently than typically developing peers (Charman et al., 2003; Pickard & Ingersoll, 2015). Therefore, individuals with ASD may be disadvantaged when receptive language scores are contingent on gesture use.

Some receptive language tasks in the PLS-5 require examinees to use eye contact and to establish communication for a social purpose (Volden et al., 2011). For example, one pass/fail item examines whether the child looks intently at a speaker. These tasks could be particularly difficult for individuals with ASD because individuals with ASD use less eye contact and initiate and respond to social communication less frequently than typically developing peers (Pickard & Ingersoll, 2015). Therefore, it is difficult to determine if low scores on receptive language measures reflect true receptive language deficits, rather than social communication deficits.

Issues such as social unresponsiveness, environmental distractions, unfamiliarity of the testing environment, lack of a pointing response, inattention, perseverative responding, anxiety, and frustration as test items get more difficult affect the reliability and validity of direct receptive language assessments (Plesa Skwerer et al., 2016). Many children may exhibit these behaviors

during testing; but they are more common among individuals with ASD (Prelock & McCauley, 2012, Chapter 6). In order to obtain valid information on receptive language skills, it is important to differentiate between social and behavioral deficits and language comprehension deficits.

**Indirect Assessments.** Indirect assessments, such as parent or caregiver reports, can alleviate the behavioral issues that occur during direct assessment, and provide information about the child's functioning in multiple natural contexts. These instruments are also easy to distribute, and are less time-consuming than direct assessments. Unfortunately, there are some validity concerns regarding indirect assessments; along with concerns regarding their appropriateness for individuals with ASD.

**Validity issues.** When filling out indirect assessments, parents and caregivers infer comprehension based on their child's behavior (Brady et al., 2014). Parents and caregivers are often not language experts and do not have the skills necessary to determine language comprehension. Tomasello and Mervis (1994) describe three situations when parents may incorrectly assume language comprehension based on their child's behavior. First, parents may assume comprehension when their child looks up in response to a spoken word. The child may be responding to their parent's voice but may not understand the meaning of the spoken word. Second, parents may assume comprehension when a child obeys a complex verbal demand that is embedded in a familiar routine and paired with gestures. For example, a child may clean up after a parent points to their toy bin and says "time to clean up your toys." The child may know that they put their toys away every night before bed, and when their parent points to their toy bin, they need to put their toys away. The child may be responding to a familiar routine or their parent's gestures rather than comprehending the phrase "time to clean up your toys." Finally, parents may misinterpret familiarity with an object as comprehension of the object's label

(Tomasello & Mervis, 1994). For example, a child may often play with toy cars; however, this does not indicate comprehension of the word *car*.

Researchers have also questioned the concurrent validity between indirect and direct receptive language assessments. Concurrent validity examines the agreement between a person's outcomes on two different assessments that measure the same construct. In a 2003 study, Charman and colleagues found good agreement between direct and indirect measures of expressive language, but poor agreement between direct and indirect measures of receptive language. Other studies have reported good agreement between direct and indirect measures for both expressive and receptive language components; however, there is often higher agreement on expressive subtests than receptive subtests (Hudry et al., 2010; R. J. Luyster et al., 2008).

It is easy for parents to overestimate their child's receptive language abilities, resulting in inaccurate outcomes on indirect measures. Potentially overinflated norms on indirect measures, such as the MacArthur-Bates Communicative Development Inventories (MCDI; Fenson et al., 1993), may make individuals with ASD look even more delayed in comparison to typically developing children (Tomasello & Mervis, 1994). Tomasello and Mervis (1994) found that word comprehension scores decreased when the MCDI was administered in an interview format where word comprehension is briefly defined before the subtest is administered, as opposed to when parents complete the MCDI independently. However, MCDI norms were created from assessments that parents completed independently; yet, studies that include children with ASD often administer the MCDI in an interview format (Hudry et al., 2010; R. J. Luyster et al., 2008). Thus, results from these studies may depict an inaccurate and exaggerated picture of receptive language delays in children with ASD compared to peers with typical development.

***ASD-related bias.*** The MCDI is often used in research and clinical practice for school-aged children with ASD who are minimally verbal (Kasari et al., 2013); however, age differences between children with ASD and sample norms can cause bias in certain test items (Bruckner et al., 2007). Bruckner, Yoder, Stone, and Saylor (2007) found that 25 items in the receptive vocabulary checklist on the MCDI show a large bias and should not be used when assessing children with ASD who are older than 16-months. Some of these items include hello, bottle, the child's name, peek-a-boo, stroller, kitty, banana, and keys. It may be difficult for parents to judge comprehension of words, such as bottle, that are not prevalent in older children's environments.

Bruckner and colleagues (2007) also noted that characteristics such as orienting deficits, social communication deficits, and restricted object use distinguish children with ASD from developmentally-matched peers with typical development and can pose challenges for parents to interpret language comprehension in specific contexts (Bruckner et al., 2007). Therefore, parents' responses to orientation and joint attention questions, such as these from the Receptive Language subtest on the Vineland Adaptive Behavior Scales-II (Vineland; Sparrow, Cicchetti, & Balla, 2005): "Looks toward parent or caregiver when hearing parent's or caregiver's voice." and "Responds to his or her name spoken (for example, turns toward speaker, smiles, etc.)." may show orientation or social communication deficits rather than receptive language deficits. Thus, items that focus on eye contact, joint attention, and words that are not relevant to the child's environment, can have a large impact on the validity of assessment outcomes for individuals with ASD.

### **Alternative Receptive Language Modalities and Procedures**

Current research shows alternative assessment modalities and procedures may yield more accurate assessment results for this population than traditional receptive language assessments (Gerhard et al., 2016; Plesa Skwerer et al., 2016).

**Touchscreen Assessments.** Children with ASD typically have fine motor difficulties (Dawson & Watling, 2000) that may make pencil-and-paper assessments challenging. Making a direct selection on a computer, tablet, or iPad may be easier and more efficient for individuals with ASD than using a writing utensil to make a selection on paper (Bouck, Savage, Meyer, Taber-Doughty, & Hunley, 2014). In 2014 Bouck and colleagues conducted an intervention study with three students with ASD, two who had verbal skills and one who was completely nonverbal. They found that all three individuals preferred using high-tech touchscreens than low-tech pencil/paper to complete task demands; as a result, researchers concluded that students with ASD might maintain higher levels of engagement using high-tech devices opposed to low-tech devices (Bouck et al., 2014). Consequently, individuals with ASD may attend for longer periods of time during touchscreen assessments than traditional low-tech assessments.

Plesa Skwerer and colleagues (2016) conducted one of the first studies comparing the use of high-tech alternative modalities as measures of receptive language skills in individuals with ASD who are minimally verbal. This study compared participants' scores across many adapted measurement tools, including a modified PPVT-4, a touchscreen assessment, and an eye-tracking assessment. Plesa Skwerer and colleagues (2016) claimed that some participants who were untestable on the PPVT were not only able to complete the touchscreen assessment, but scored above 75% accuracy on this measure. However, their touchscreen and eye-tracking assessments included additional words that are not assessed in the PPVT. Similarly, their eye-gaze and touchscreen trials consisted of one target word and one foil word, as opposed to PPVT trials that



consist of one target word and three foil words. Therefore, differences in participants' scores across assessments may have been caused by variables other than assessment modality (i.e., using different target words and having fewer foil items).

**Real-Object Stimuli.** Pictures are abstract representations of real objects. Over time, infants learn the concept of a picture and understand that pictures represent objects (DeLoache, Pierroutsakos, Uttal, Rosengren, & Gottlieb, 1998). Individuals with ASD who are minimally verbal have difficulty understanding the symbolic meaning of pictures (Preissler, 2008). Therefore, individuals with ASD may perform better on receptive language assessments that use real-object stimuli as opposed to picture stimuli. Cocking and McHale (1981) found that children with typical development performed better on language assessment tasks that used object stimuli than on identical tasks that used picture stimuli.

In addition, recent research with infants and adults has shown attentional and information-processing advantages for real-object stimuli as opposed to picture stimuli. Infants can differentiate between known and novel objects more quickly with real-object stimuli than with picture stimuli (Carver et al., 2006; Gerhard et al., 2016). Infants also prefer looking at real-objects as opposed to a picture of the same object (Gerhard et al., 2016). Similarly, Gomez, Skiba, and Snow (2018) found that adults spend longer periods of time attending to real-object stimuli than to picture stimuli. Real-objects have affordance, meaning real-objects have potential for physical action, which may impact attention. Additionally, three-dimensional objects are more salient and visually richer than pictures; and thus, word-object matching may be easier than word-picture matching.

### **Alternative Receptive Language Assessments**

**Eye Gaze.** In attempts to alleviate difficult response demands for individuals with ASD, such as pointing to pictures, recent research has begun to look at more implicit measures of receptive language. Implicit measures, such as eye gaze, do not require comprehension of test instructions or an outward motor response (Brady et al., 2014). The intermodal preferential looking paradigm (IPLP; Golinkoff, Hirsch-Pasek, Cauley, & Gordon, 1987) is a measure of comprehension that uses eye gaze. In IPLP, examinees are shown two visual stimuli at the same time, one stimulus in the left field of vision and one in the right field of vision, then they hear an auditory stimulus that corresponds with one of the visual stimuli (Golinkoff et al., 1987). The examinees' responses are videotaped and later analyzed using a frame-by-frame analysis procedure to determine the proportion of time spent looking at each stimulus. If the examinee spends more time looking at the visual stimulus that matches the auditory stimulus, then they are judged to have understanding of the word.

Infant outcomes on IPLP have high agreement with parent-reports on word comprehension (Robinson, Shore, Hull Smith, & Martinelli, 2000). IPLP has also been used to assess children with ASD (Goodwin, Fein, & Naigles, 2012; Swensen, Kelley, Fein, & Naigles, 2007). Using IPLP, Swensen and colleagues (2007) were able to measure emergent language comprehension in children with ASD. They found that children with ASD were able to comprehend sentences in specific grammatical structures, before they could produce these sentence structures (Swensen et al., 2007). Unlike findings from direct standardized measures (Ellis Weismer et al., 2010; Hudry et al., 2010; Volden et al., 2011), Swenson and colleagues' (2007) findings suggest that individuals with ASD may have language profiles that are similar to peers with typical development, where receptive skills are more advanced than expressive skills. IPLP may be a more sensitive receptive language measure for individuals with ASD than current

direct standardized measures that require more working memory and overt motor movements. Unfortunately, the frame-by-frame video coding that is necessary for IPLP analysis can be very time consuming.

Eye tracking technology provides more thorough and precise data on eye gaze patterns; and it allows researchers to obtain and analyze data more efficiently than IPLP. Some recent studies have shown that eye tracking technology is a valid measure of word comprehension in children with ASD (Bavin et al., 2014; Brady et al; 2014; Venker et al., 2013). Brady and colleagues (2014) used the PPVT to identify sets of known and unknown words in children with ASD and children with typical development. The children were then tested on the same sets of words using eye tracking technology. Both children with ASD and children with typical development spent a significantly longer amount of time looking at target pictures for known words. Children with ASD spent equal amounts of time looking at target and non-target pictures when the word was unknown. These findings suggest that eye tracking is a valid measure of word comprehension in children with ASD and with children with typical development. They also suggest that eye tracking is a potentially viable alternative measure of word comprehension for children with ASD who score at the floor of, or are unable to complete other measures (Brady et al., 2014).

Plesa Skwerer and colleagues (2016) used using eye tracking to measure word comprehension in older children and teens with ASD who had minimal verbal skills. Unlike in Brady and colleagues' (2014) study where participants were assessed using the same words on both standardized and eye tracking measures, Plesa Skwerer and colleagues (2016) assessed word comprehension skills using direct (PPVT) measures, but used a different set of 84 words to assess word comprehension using an eye tracking measure and a touchscreen measure. In the eye

tracking assessment, participants looked at a computer screen that had an eye-tracking camera embedded in the computer display. In each trial, two pictures appeared on the computer screen while a voice recording played that labeled the target word (e.g., "Look [target word]").

Researchers found that participants looked proportionally longer at target pictures than at non-target pictures, suggesting that eye tracking measures can provide evidence of word comprehension for children and adolescents who are minimally verbal and have ASD (Plesa Skwerer et al., 2016). Eye tracking data also provided insight about how spoken words influence attention monitoring in children and teens with ASD who are minimally verbal. Researchers assessed an overall measure of attention allocation to the stimuli, by examining the number of trials in which the participants looked at either stimuli (target or non-target) after hearing the word. Trials in which participants did not orient to either stimuli, post-word onset, were considered unreliable and were not used in data analysis. One participant failed calibration and was unable to complete the eye tracking measure. Of the remaining 18 participants, contributions of usable data trials ranged from 36% to 97%, with a group mean of 67% (Plesa Skwerer et al., 2016). Overall, about 48% of eye gaze trials were lost or had to be removed from statistical analyses (Plesa Skwerer et al., 2016). According to Plesa Skwerer and colleagues (2016), the substantial variation of reliable eye gaze trials suggests that individuals with ASD who have minimal verbal skills have heterogeneous attentional skills.

Despite noted limitations, eye gaze technology is still a promising new alternative measurement of receptive language comprehension for individuals with ASD who are nonverbal. These assessments place fewer demands on working memory, comprehension of test instructions, and motor responding; and thus, may be more sensitive measures of receptive language skills in individuals with ASD than traditional language assessments. However more

research is needed before eye tracking can be used to measure comprehension in clinical practice.

## **Current Study**

Further research is needed to determine whether testing modifications such as using high-tech test modalities and real-object stimuli increase performance outcomes on receptive language measures in individuals with ASD who are minimally verbal. The primary aim of this study was to evaluate and compare alternative modalities and stimuli used to measure receptive language skills in children with ASD who are minimally verbal. This study compared participants' outcomes on three different receptive language assessments: a standardized assessment, a touchscreen assessment, and an assessment that uses object-stimuli as opposed to picture-stimuli.

The current study controlled for the task demand and items across all three assessments in order to identify how differences in assessment condition impact performance on receptive vocabulary assessments. Controlling for the items across conditions means that all three tasks used the same target and foil words and required the same response. The only differences across conditions was the modality (i.e., how stimuli are presented). Maintaining experimental control across all three assessment tasks was necessary in order to verify that any difference in performance between assessments was caused by the assessment rather than extraneous variables.

Previous research suggests that children and infants have information processing advantages for objects as opposed to pictures (Gerhard et al., 2016). Objects are more tangible and might be more salient than pictures. Gerhard and colleagues also found that infants prefer looking at objects than pictures (2016). Therefore, the real-object assessment may require less

cognitive-load and might be more visually appealing than assessment conditions that use pictures. Thus, I hypothesized that participants would score higher on the real-object assessment than on assessment conditions that use pictures.

I also hypothesized that participants would earn better scores on the touchscreen assessment than on the low-tech assessment. Previous research found that individuals with ASD with minimal and more advanced verbal skills, prefer to use high-tech touchscreen devices rather than low-tech paper and pencil to complete work-like tasks (Bouck et al., 2014). Individuals with IDD exhibit higher levels of attention maintenance and lower rates of problem behavior during preferred tasks compared to non-preferred tasks (Vaughn & Horner, 1997).

A secondary aim of this study was to examine how individual characteristics such as cognition, maladaptive behavior, autism symptomology, and gesture use impact performance on assessment conditions. I hypothesized that participants with higher scores on cognitive tests would score higher across all PPVT assessments than participants with low cognitive scores. Additionally, I hypothesized that parental reports of high rates of maladaptive behavior would be associated with lower scores across all experimental assessments.

A tertiary aim of this study involved the use of IPLP to determine whether eye gaze data could be used to measure implicit word comprehension for children with ASD. Recent research shows that children with typical development spent more time looking at target than non-target pictures, when previous testing showed the target word was unknown (Brady et al., 2014). It is possible that some of these words were emerging in the child's lexicon, but they were not yet mastered. Examining participants' eye gaze data could provide valuable information on mastered

and emerging word-knowledge, as well as information on attention and processing skills in this population.

Finally, I also summarized clinical factors associated with each assessment condition. Specifically, I described assessment duration and occurrence of problem behavior in each assessment condition. These factors influence the ease of administration and clinical feasibility of each assessment condition.

## **Method**

### **Participants**

Participants were twenty-seven children with ASD who had minimal verbal skills, which was defined as 30 or fewer functional words. Participants' hearing, and vision were within normal limits or were corrected to be within normal limits. Parental report was relied upon to determine participants' ASD diagnosis, number of spoken words, hearing, and vision. However, the Childhood Autism Rating Scale-Second Edition (CARS) was administered to confirm ASD diagnosis, and participants who produced more than 30 words during classroom observations or experimental assessments were excluded from the study.

The majority of participants lived in the Kansas City metropolitan area. Participants were between the ages of 3 years 8 months and 12 years 2 months ( $M = 7$  years and 7 months old;  $SD = 30.77$  months). Four participants were females and the remaining 23 participants were males. Twenty-four participants were Caucasian, five were Black or African American, and one participant was Asian. Five participants were Hispanic or Latinx. Table 1 presents the descriptive characteristics of the participants. Participants' familiarity with the experimenter varied. Thirteen participants were unfamiliar to the experimenter prior to the assessments, 10 participants had

seen the experimenter observing or working in his or her school more than once prior to the assessments, and four participants had worked with the examiner prior to the assessments.

Participants were recruited via school districts, speech-language pathology private practices, and behavioral clinics in Topeka, Kansas and in the greater Kansas City area. Recruitment information and consent forms were disseminated by teachers, speech-language-pathologists, and other professionals who worked with children who met qualification criteria. After receiving written consent from the parent or guardian, the experimenter contacted the caregiver to conduct initial screeners to ensure the child met inclusionary criteria. Upon completion of all assessments, participants were given \$20 as compensation for their time. All methods were approved by the Institutional Review Board at the University of Kansas.

### **Diagnostic, Cognitive, and Behavioral Measures**

**Childhood Autism Rating Scale-Second Edition** (CARS-2; Schopler, Van Bourgondien, Wellman, & Love, 2010). The CARS was used to verify participants' ASD diagnosis and to quantify ASD symptomology. The CARS is a direct observation measure with a 4-point rating scale that is used to describe the frequency, duration, intensity, and idiosyncrasy of behaviors. A score of 30 or higher on this assessment is indicative of ASD, and thus a score of 30 or higher was required to participate in this study. In a recent systematic review, McConachie and colleagues (2015) examined the measurement properties and qualities of tools used to measure ASD, and concluded the CARS was one of the most valid available measures.

The CARS was administered by the experimenter after observing the child in his or her natural home or school environment and upon completion of experimental and MSEL assessments. Observation duration varied among participants, but in most cases a minimum of 1.5 hours of observation was needed to answer all the questions. If specific situations or



behaviors were not observed, caregivers, teachers, or other professionals who were familiar with the participant were asked questions to help determine an appropriate rating.

**Mullen Scales of Early Learning (MSEL;** Mullen, 1995). Participants Visual Reception, Receptive Language, Expressive Language, and Fine Motor skills were assessed using the MSEL. The MSEL is a direct assessment used to measure early learning in children from birth to 6 years and 8 months of age. However, this assessment is often used to assess individuals with IDD who are older than the normative sample (Farmer, Golden, & Thurm, 2016). The test manual reports high internal consistency for the MSEL subtests, with median reliability ranging from .75 to .83 (Mullen, 1995). Correlations between the MSEL and Bayley Scales of Infant Development (range of .53 to .59) support the validity of the MSEL as a measure of cognitive ability for infants and children up to 6 years and 8 months of age (Mullen, 1995).

The majority of MSEL assessments were administered by the experimenter within a month of participant's experimental assessment date. Three participants had previously completed the MSEL assessment with a reliable administrator within three months of his or her experimental assessment date, so those scores were used for this study. The majority of participants completed all of the MSEL subtests in the same day. However, due to various reasons such as scheduling constraints and testing fatigue, some participants completed the MSEL on two separate days. The MSEL took approximately 25 minutes to an hour to complete.

**Caregiver Questionnaire.** The caregiver questionnaire included the Maladaptive Behavior Index from the Vineland Adaptive Behavior Scales-II (Vineland; Sparrow, Cicchetti, & Balla, 2005) and 12 questions regarding gesture use and history with touchscreen devices (see Appendix A). The Maladaptive Behavior Index consists of 50 questions that require caregivers to rate the frequency of their child's internalizing, externalizing, and other problem behavior. It is

appropriate for use with individuals from birth to adulthood and has been normed across different disability, including individuals with autism who are nonverbal. The Vineland test manual reports high internal consistency (correlations ranging from .85 to .91 across age groups) and high test-retest reliability (correlations ranging from .74 to .98) for the Maladaptive Behavior Index (Sparrow et al., 2005). Concurrent validity was also examined by comparing scores on the Maladaptive Behavior Index to scores on the Behavior Symptoms Index of the Behavior Assessment System for Children, Second Edition (BASC-2). Scores on both subtests were highly correlated, .80 (Sparrow et al., 2005).

The additional 12 gesture and touchscreen questions were created by the experimenter. Eight questions asked caregivers to rate their child's ability to produce early and conventional gestures and four multiple-choice and rating scale questions asked caregivers to describe their child's touchscreen usage and preference. The caregiver assessment provided quantifiable scores to describe participants' maladaptive behaviors, gesture use, touchscreen history and touchscreen preference.

Caregivers had the option of completing the caregiver questionnaire in person, over the phone, or having the questions sent via email. Caregiver questionnaires were conducted by the experimenter and took approximately 15-30 minutes to complete. The majority of caregiver questionnaires were completed within a month of the experimental assessments; however, some questionnaires were completed up to 6 months after the experimental assessments were conducted.

Table 1. *Participant Characteristics*

Characteristic	<i>N</i>	Mean	<i>SD</i>	Range
MSEL: Visual Reception	27	29.37	6.42	16 - 46
MSEL: Fine Motor	27	26.19	5.39	17 - 38
MSEL: Receptive Language	27	22.70	6.60	11 - 38
MSEL: Expressive Language	27	16.70	6.14	9 - 33
CARS: ASD Symptomology	27	41.13	4.44	30.5 - 50
Vineland: Maladaptive Behavior	24	32.96	9.04	19 - 47

Note. MSEL = Mullen Scales of Early Learning; CARS = Childhood Autism Rating Scale; Vineland = Vineland Adaptive Behavior Scales. All assessment scores are raw scores.

### Experimental Assessment Conditions

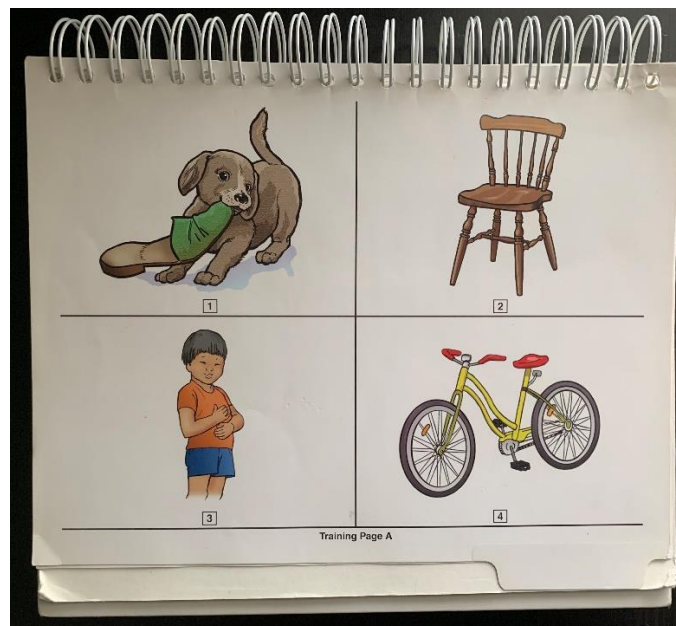
This study included three experimental assessments that were modified versions of the standardized PPVT-4: the standardized PPVT with modified administration, a touchscreen assessment, and an assessment using object stimuli. All three assessments used the same number of trials and had the same target and foil words in each trial. Experimental assessments were also identical with regard to administration and task demands (i.e., identifying the picture/object that matches the auditory stimulus). The differences across experimental procedures were the modality in which the stimuli were presented (i.e., low-tech paper, high-tech touchscreen, or 3D object) and/or the stimuli themselves (i.e., picture or object).

**Standardized Assessment Condition: PPVT-4** (Dunn & Dunn, 1997). The PPVT measures receptive vocabulary skills in individuals from two-and-a-half years old up to 90+ years. In this standardized assessment, examinees are shown a page that is divided into four quadrants, with four pictures (one target and three foil) in the center of each quadrant. Examinees are asked to point to the target word after it is named (e.g., "point to *boy*"). The assessment begins with four training items that are designed to teach the individual how to give a desired response (i.e., pointing to or touching the correct picture). During each training trial, examiners provide descriptive feedback on response accuracy. Typically, if an individual is unable to give a correct response in at least two out of the four training items, the assessment is discontinued. In

standardized administration, test items are administered after the examinee passes at least two of the training trials. The examiner administers the first test item that corresponds with the examinee's age and may administer test items for younger age groups to establish a basal.

However, for this study, administration was modified so that all four training trials were administered for each participant. Participants then completed test questions 1-12 regardless of their age or performance on the training trials. The assessment ended after question 12 even if a ceiling was not reached. Figure 1 shows the visual stimuli used in the training trials of this assessment.

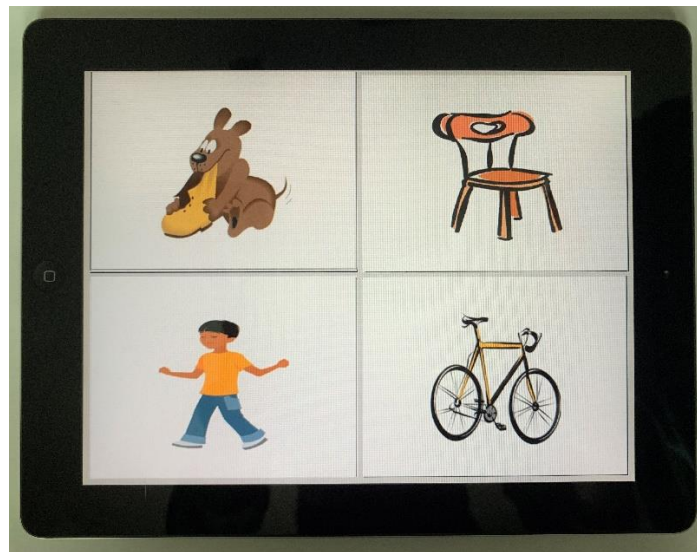
Figure 1. Training trials for the standardized assessment condition.



**Touchscreen Assessment Condition.** The touchscreen task was administered using Paradigm software (2016 Perception Research Systems Incorporated) on an iPad. Paradigm is a stimulus presentation system that allows researchers to create experiments on computer and mobile devices. Paradigm allows researchers to create interactive assessment and intervention experiments with customizable stimuli and layout. The touchscreen version of the PPVT used the same sets of pictures, displayed in the same quadrant-format as the standardized booklet-format

PPVT. However, due to copyright reasons, the pictures were not identical to the ones used in the standardized PPVT. The pictures used for the touchscreen assessment condition were the same picture stimuli that Brady and colleagues used in their 2014 eye gaze study (Brady et al., 2014). Figure 2 shows the visual stimuli for the training trials of the touchscreen assessment.

Figure 2. Training trials for the touchscreen assessment condition.



**Real-Object Assessment Condition.** For this assessment, real objects, such as a toy ball, a plastic figure of a dog, and a toy chair were used as stimuli. Dolls and objects were glued together for each stimulus in the verb trial. For example, a toy apple slice was glued to the mouth and hand of a doll to represent the target “eating,” and a doll was glued to a swing to represent the foil “swinging.” Visual stimuli were presented in a straight line on a piece of cardboard on the table in front of the participant, close enough for the participant to touch each object. The cardboard had four small dots with equal distance between dots. Objects were placed on these dots to ensure uniform placement of objects across trials and across participants. Each real-object assessment condition was recorded using three GoPro cameras that were placed directly behind the stimuli. Each camera was placed in a specific location to capture participants’ eye gaze and

the stimuli from three different angles. Figure 3 shows the layout of the real-object assessment and the visual stimuli for the training trials of this assessment.

Figure 3. Training trials, camera placement, and stimuli placement for the real-object assessment condition.



## Procedures

After obtaining written consent from guardians, initial screenings were conducted with potential participants' parent, caregiver, or teacher to ensure participants met inclusionary criteria as well as to gain information to help prepare for the assessment (see Appendix B). Caregivers or teachers were asked about the child's attention span, preferred items and activities that could be used as rewards during the assessment, and challenging behaviors.

The PPVT experimental assessment conditions (i.e., standardized assessment condition, touchscreen assessment condition, and real-object assessment condition) were administered in a single day by the experimenter. A few participants completed the experimental assessment conditions on the same day as the MSEL and CARS assessments, but most participants completed the MSEL and CARS assessments on a separate day than the experimental assessments. The three experimental assessment conditions took a total of approximately 20-40 minutes to administer. Assessments were conducted one-on-one or with a parent,

paraprofessional, teacher, or behavior specialist sitting close by. The experimenter explained the assessment procedures to adults and asked them to refrain from prompting the child's responses in any way throughout the assessments. If needed, adults helped manage problem behavior that occurred during the assessments.

Assessment order was randomized to control for order effects. With three conditions (A: standardized, B: touchscreen, C: real-object) there was a total of six possible presentation orders (i.e., ABC, CAB, BCA, ACB, BAC, CBA). Twenty-four participants were randomly placed into the six presentation order groups, so that each group contained four participants. The experimenter considered excluding participants who earned perfect scores on all three assessment conditions from the final analysis, so the final three participants were intentionally placed in presentation order groups that contained participants who earned perfect scores across all three assessment conditions. The experimenter later chose to include participants who earned perfect scores in the final analysis, which resulted in unequal numbers of participants in presentation groups (4 participants: ABC, CAB, BAC, CBA; 5 participants: BCA; 6 participants: ACB).

**Experimental Assessments.** Assessments were conducted in participants' homes, schools, behavioral clinics, and one was conducted in a study room at a public library. Before testing began, the experimenter spent approximately 5 minutes building rapport with the participant and orienting them to the testing environment. Three GoPro cameras were placed on the table to record the participants eye gaze and responses during the real-object assessment (see Figure 3). The examiner provided positive feedback (regardless of response accuracy) intermittently throughout the assessments to motivate the child. Verbal reinforcement was descriptive, but not related to the child's response accuracy (e.g., "Nice job sitting in your chair!")

and “I like how you’re looking at all of the pictures!”) Also, breaks, where no demands were placed on the child and the child had access to preferred items, were provided when necessary. The experimenter used clinical judgment as well as information from parents and teachers to determine when and if a participant received a break.

## **Data Collection**

**Experimental Assessments.** The experimenter collected data live, using a pencil and paper score sheet (see Appendix C). For each trial, after the stimuli were presented participants were given approximately 3 seconds to scan each stimulus before the examiner gave the task demand (“Touch X”). The examiner only gave the task demand when the participant was orienting to the stimuli or to the examiner. If the participant was not orienting towards the examiner or the stimuli, the examiner redirected the participant’s attention by saying “[Participant Name], look at all of the pictures/things” while pointing to each picture or object. If the participant touched one or more stimuli before the task demand was presented, the examiner corrected the participant by saying “wait until I tell you.” After the task demand was presented, participants were given approximately 7 seconds to respond. After the participant responded, the examiner marked the picture or item the participant selected on the data sheet.

**Multiple Responses.** If the participant touched more than one stimulus after the task demand (“Touch X”) was presented, the examiner corrected the participant by saying “only point to one” and then represented the task demand. The examiner then marked the “re-do” cell next to the appropriate trial on the data sheet (column labeled “RD”) to signify that this trial was presented twice. If the participant selected multiple stimuli more than once, the examiner marked all the stimuli the participant selected, and the response was marked as incorrect.



**No-Response.** If the participant did not respond within approximately 7 seconds after the task demand was presented, the examiner would re-present the task demand. The examiner then marked the re-do cell next to that trial on the data sheet to signify that this trial was presented twice. If the participant did not respond within approximately 7 seconds after the second presentation of the task demand, the examiner marked the “no-response” cell next to the appropriate trial on the data sheet (column labeled “NR”) to signify that the participant did not respond to this trial.

**Eye-Tracking.** The experimenter and a trained research assistant observed participants' eye-gaze from videos taken during the real-object assessment condition. The video from the camera in the center position (see Figure 3) was used for eye gaze coding. Videos taken from the left or right camera position were referenced during the rare occurrence when coders could not see the participants' eye gaze from the center video. Videos were played on VLC Media Player, and playback speed controls were used to reduce the speed by 0.3x times the normal speed. Coders downloaded the VLC Time Extension which displayed the video runtime in hours, minutes, seconds, and milliseconds.

Eye gaze was coded during the response period of each trial. The response period started immediately after the demand was presented (“Touch X”) and the response period ended once the participant selected a stimulus. For no-response trials, the response period ended when the examiner removed the stimuli. Coders noted the time participants spent looking at 1) the target stimulus, 2) the foil stimuli, and 3) non-stimuli (i.e., looking at anything other than the stimuli). Coders paused the video each time a participants' eye gaze shifted, and marked the “start time,” when participants started looking at a stimulus, and the “end time,” when his or her eyes shifted away from that stimulus. For each trial, total number of milliseconds spent looking at the target

stimulus, foil stimuli, and non-stimuli were calculated by subtracting the start time from the stop time for each fixation period. The total time spent looking at foil items was divided by 3 to get an average looking time at foil stimuli. Average foil looking time was calculated to adjust for the relative difference in number of target versus number of foil stimuli (target = 1 of 4, and foil = 3 of 4).

**Problem Behavior & Assessment Duration.** Appendix D includes a list of operational definitions for each problem behavior topography (e.g., aggression, self-injury, elopement) that were tracked during assessments. The experimenter recorded whether problem behavior, or attempts, occurred during each trial of each assessment by marking the “problem behavior” cell (in the column labeled “PB”) next to the appropriate trial on the data sheet (see Appendix C). A stopwatch was used to record the duration of each assessment. Training items and any breaks were included in the total assessment duration.

**Inter-Rater Reliability.** An undergraduate student was trained to collect reliability data on participants’ responses and occurrence of problem behavior for all experimental assessments. The research assistant was given operational definitions and examples of each type of participant response and each type of problem behavior that was tracked. Then the research assistant and examiner watched a video of all three assessment conditions and scored each assessment together.

The research assistant independently collected data, on 30% of standardized PPVT, touchscreen, and real-object assessments. Reliability data was collected live and from videos of previously recorded assessments. Agreement was calculated using point-by-point agreement for each assessment. The number of agreements for each assessment was divided by the total number of trials for that assessment and this number was multiplied by 100. Table 2 shows the

percent agreement for stimulus selection, problem behavior, and no-response trials for all three assessment conditions.

***Eye-Gaze Reliability.*** The same research assistant was also trained to collect eye-gaze data from real-object assessment condition videos. During eye-gaze training the examiner and research assistant watched trials together and talked through coding procedures. Then the research assistant and examiner coded videos independently and met to talk about disagreements. Trials with disagreements were discussed, and consensus coded.

After training, the research assistant was the primary eye-gaze coder. The examiner coded 20% of the trials used for Aim 3 analysis. Primary and reliability coders' times were compared by examining the total time spent looking in each looking category (i.e., looking at target, looking at foils, and looking at non-stimuli) for each trial. Percent agreement for each looking category in each trial was calculated by dividing the shorter time by the longer time and multiplying this number by 100. Then, percent agreement for each trial was calculated by averaging the percent agreement of the three looking categories. Overall agreement was calculated by adding the percent agreement for each trial, dividing this number by the total number of trials, and multiplying this number by 100. Total agreement between primary and reliability eye-gaze coders was 76.4%.

*Table 2. Percent Agreement of Participants' Responses and Behavior*

Variable	Experimental Assessment Condition		
	Standardized	Touchscreen	Real-Object
Stimulus Selection	98.96%	96.88%	98.96%
Problem Behavior	96.88%	93.75%	98.96%
No-Response	100%	100%	100%

**Fidelity.** A trained research assistant used a fidelity checklist (see Appendix E) to check administration fidelity in 22% of administrations for each experimental assessment condition. Fidelity checks were conducted in-person ( $n = 4$ ) and by watching assessment videos ( $n = 2$ ). Assessment administrations that were checked for fidelity were chosen before participants were recruited, by randomly selecting six of the 27 participant identification numbers. Participants were assigned identification numbers in chronological order by date their consent form was received. To determine a fidelity score, the number of steps correctly implemented during the assessment administration was divided by the number of steps implemented, and then multiplied by 100. All PPVT experimental assessments were administered with high fidelity: standardized assessment = 99.65%, touchscreen assessment = 98.96%, and real-object = 97.92%.

### **Data Analysis**

The primary aim of this study was to evaluate the use of different receptive language assessment modalities and stimuli for children who are minimally verbal, by comparing outcomes between different assessment conditions. A crossed random effects model was used to predict variability across participants and across experimental assessment items. Individual responses to each experimental assessment item at level 1 were nested within participants, within experimental assessment condition, and within items. Sequential models with item predictors (experimental assessment condition and presentation order) were tested to examine mean and variance difference across experimental assessment conditions.

The second aim of this study was to describe how participant characteristics impact assessment performance. Participant predictors (i.e., MSEL scores, Vineland scores, and CARS scores) were added to the crossed-random effects model to account for participant variance.

These additional models estimated the degree to which individual characteristics were responsible for variance in scores.

Crossed random effects models were used instead of a traditional Analysis of Variance (ANOVA) because crossed random effects models include all random item variation *and* all random participant variation in the model. ANOVA models, however, use aggregate data for items or participants. Treating items or participants as fixed when they are not can increase the likelihood of making a Type 1 error when testing the effects of predictors. Accounting for multiple sources of variation allows for the most accurate testing of predictor effects (Hoffman, 2015). In addition, crossed random effects models can include continuous and categorical variables, rather than just categorical variables.

Descriptive statistics were used to describe the efficiency and ease of assessment for each assessment condition. Ease and efficiency of assessment were examined by looking at the average duration of each assessment, the number of no-response trials in each assessment, as well as the occurrence of problem behavior in each assessment. Assessment duration, problem behavior, and no-response trials were also compared across conditions using repeated-measures ANOVAs. Post-hoc Bonferroni procedures were conducted to examine pairwise differences between conditions. Alpha levels were adjusted for post-hoc pairwise comparisons to account for the number of comparisons made. Thus, instead of using .05 alpha levels, an adjusted alpha level of .0167 was used ( $.05/3$ ). The sum of trials with problem behavior were calculated for each participant and each assessment by adding the number of trials where the participant engaged in problem behavior in each assessment. The average duration and range of assessment duration were reported for each experimental assessment condition. No-response trials were measured by

summing the total number of trials where participants did not respond after the second presentation of the stimulus.

The third and final aim of this study involved the use of IPLP to determine whether eye gaze data may provide additional, implicit information about word comprehension for children with ASD. Trials where participants responded correctly via overt responses (i.e., touching the correct item) were not included in this analysis. Eye-gaze data from each incorrect response in the real-object assessment were examined to gain implicit information about word-understanding. A paired-samples t-test was conducted to compare the time spent looking at target stimuli and the average time spent looking at foil stimuli (total foil looking time divided by 3) during incorrect trials. A repeated measures ANOVA was also used to examine time spent looking at target stimuli, average time spent looking at foil stimuli, *and* time spent looking at non-stimuli during each incorrect trial. A post-hoc Bonferroni procedure, with adjusted alpha levels of .0167, was conducted to examine differences in looking-time between each group (i.e., target stimuli, foil stimuli, and non-stimuli). Descriptive statistics were then used to describe eye-gaze during no-response trials.

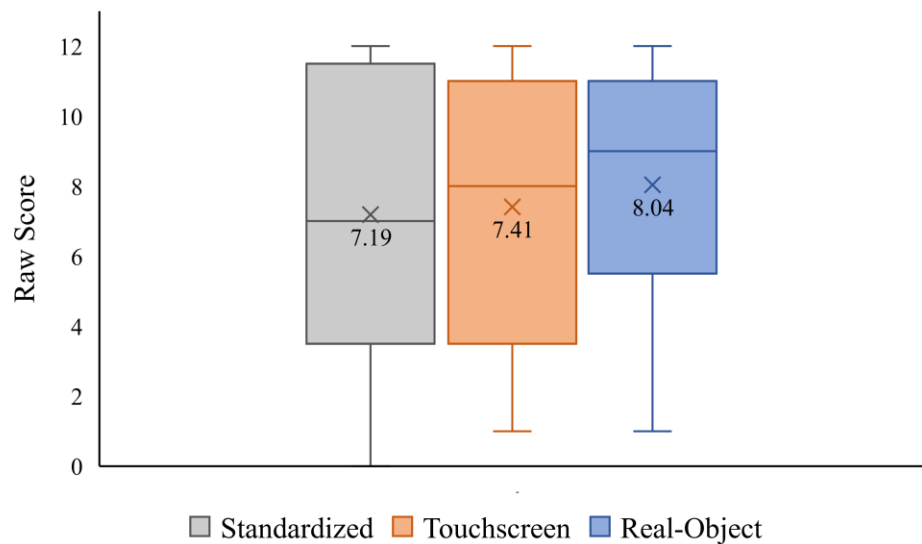
## **Results**

### **Performance on Assessment Conditions**

Figure 4 displays a box plot with descriptive statistics of participants' raw scores for each assessment condition. The mean score was higher in the real-object assessment condition ( $M = 8.04$ ;  $SD = 3.65$ ) than in the touchscreen ( $M = 7.41$ ;  $SD = 3.83$ ) or standardized ( $M = 7.19$ ;  $SD = 4.45$ ) assessment conditions. Four participants earned perfect scores on all three assessment conditions; and five participants scored at or below chance level (25% accuracy) on all three

assessment conditions. Eight participants earned better scores on the real-object assessment condition than in the standardized and touchscreen conditions. Four participants earned better scores on the touchscreen assessment condition than on the real-object and standardized conditions. Four participants received their highest score in the standardized condition, and seven participants earned their highest score in more than one condition. Twelve participants received their lowest score on the standardized assessment condition.

Figure 4. Box Plot of Descriptive Statistics in Experimental Assessment Conditions



A crossed random effects model was used to predict variability in responding across participants, across items, and across assessment conditions. Twenty-seven participants responded to 12 items in three assessment conditions (972 total items). Restricted maximum likelihood was used to estimate the models, where individual responses to each experimental assessment item at level 1 were nested within participants and within items at level 2. Significance of fixed effects (means) were evaluated using Wald test  $p$ -values, and significance of random effects (variance) were evaluated using  $-2\Delta LL$  tests (see Hoffman, 2015).

First, a model without predictors (i.e., empty means model) was calculated to estimate the probability of answering correctly across all items. This model was used as a baseline model. Then, additional models were estimated to partition the variance in responding between participants and assessment items. There was significant variability in responding across subjects,  $-2\Delta LL(1) = 368, p < .001$ , and across items,  $-2\Delta LL(1) = 67, p < .001$ . Across all items ( $M = 0.87, SE = 0.42, N = 12$ ), ball  $t(934) = 3.43, p < .001$ , and banana  $t(934) = 2.34, p = .0196$ , were significantly easier than average items, and spoon  $t(934) = -3.25, p = .001$ , and mouth  $t(934) = -3.81, p < .001$ , were significantly more difficult than average items. The person variance in probability of answering correctly was 5.55 and the item easiness variance was 0.87. Table 3 displays the ranking of items by difficulty from easiest to hardest. Both subject and item variation were significant, thus their random effects were retained in the model.

Table 3. Ranking of Experimental Assessment Items from Easiest to Hardest

Target Word	Rank	Estimated Mean Difference	Standard Error	df	t Value	Significance
Ball	1	*1.4542	0.4241	934	3.43	0.0006
Banana	2	*0.9502	0.4065	934	2.34	0.0196
Bus	3	0.6656	0.3997	934	1.66	0.0963
Dog	4	0.6656	0.3997	934	1.66	0.0963
Cup	5	0.1248	0.3916	934	0.32	0.7500
Shoe	6	0.03778	0.3909	934	0.10	0.9230
Duck	6	0.03778	0.3909	934	0.10	0.9230
Flower	7	-0.04854	0.3902	934	-0.12	0.9010
Foot	8	-0.5535	0.3893	934	-1.42	0.1555
Eating	8	-0.5535	0.3893	934	-1.42	0.1555
Spoon	9	*-1.2882	0.3968	934	-3.25	0.0012
Mouth	10	*-1.5330	0.4020	934	-3.81	0.0001

\* $p < .05$

Sequential models were tested to examine the effects of item predictors (i.e., experimental assessment condition and order). Fixed effects for assessment condition (coded such that 0 = touchscreen, 1 = standardized, and 1 = real-object), assessment order, and



assessment condition by assessment order were examined. Significant mean differences in responding were observed across assessment conditions [ $F(2, 932) = 3.58, p = 0.0282$ ]. There were no significant order effects [ $F(5, 922) = 0.56, p = 0.7277$ ] or order by condition effects [ $F(10, 922) = 1.34, p = 0.2020$ ], and thus these effects were removed.

Additional pairwise comparisons were conducted to examine differences between assessment conditions in Table 4. Real-object assessment scores were significantly higher than standardized assessment scores and were marginally higher than touchscreen assessment scores. There were no significant differences between touchscreen and standardized assessment scores.

Table 4. *Pairwise Comparisons Between Assessment Conditions*

Assessment Condition Comparison		Estimated Mean Difference	Standard Error	<i>df</i>	<i>t</i> Value	Significance
Real-Object	Standardized	*0.5933	0.2293	932	2.59	0.0098
Real-Object	Touchscreen	0.4414	0.2291	932	1.93	0.0543
Standardized	Touchscreen	-0.1518	0.2251	932	-0.67	0.5002

\* $p < .05$

**Clinical Factors.** Clinical factors associated with each assessment condition were summarized using descriptive statistics. Assessment duration, problem behavior, and no-response trials were also compared across conditions using ANOVAs. Table 5 shows descriptive statistics of assessment duration in minutes, problem behavior, and no-response trials for each condition.

No-response trials were measured by summing the total number of trials where participants did not respond after the second presentation of the stimulus. The total number of participants who had at least one no-response trial in a condition are as follows: standardized = 10, touchscreen = 5, real-object = 7. Table 5 shows the total number or no-response trials in each condition. The total number of no-response trials varied significantly among conditions ( $F(2,78)$

= 4.189,  $p = 0.0187$ ). A post-hoc Bonferroni procedure was conducted to examine pairwise differences. For pairwise comparisons, adjusted alpha levels of .0167 were used to account for the total number of comparisons (.05/3). Results showed there were significantly more no-response trials in the standardized assessment condition than in the touchscreen condition ( $t(26) = 2.833$ ,  $p = 0.0088$ ). The number of no-response trials in the real-object assessment condition was not significantly different from the number of no-response trials in the touchscreen ( $t(26) = 1.100$ ,  $p = 0.2815$ ) or standardized conditions ( $t(26) = 2.032$ ,  $p = 0.0525$ ).

Problem behavior was measured by summing the total number of trials where participants engaged in problem behavior for each assessment condition. Each of the six problem behavior topographies (i.e., elopement, tantrum, defiance, aggression, self-injury, and property destruction) occurred at least once during the experimental assessments. Occurrence of problem behavior varied among participants. Fifteen participants engaged in problem behavior during at least one trial of the standardized assessment condition, and 11 participants engaged in problem behavior during at least one trial of the touchscreen and real-object assessment. Six participants engaged in problem behavior in all three assessment conditions, and nine participants had no instances of problem behavior across all assessment conditions.

On average, participants engaged in problem behavior more frequently during the standardized condition than during the touchscreen and real-object conditions (see Table 5). However, occurrence of problem behavior did not differ significantly among assessment conditions ( $F(2,78) = 0.467$ ,  $p = 0.629$ ). Order effects were also examined to determine whether assessment order impacted the occurrence of problem behavior. There were no significant order effects on the occurrence of problem behavior ( $F(2,78) = 0.250$ ,  $p = 0.779$ ).

Assessment duration varied significantly between conditions ( $F(2,71) = 3.126, p < 0.001$ ). Post-hoc analyses were conducted using Bonferroni adjusted alpha levels of .0167 per test (.05/3). Results revealed that the real-object assessment condition took much longer to administer than the standardized ( $t(25) = -13.021, p < 0.001$ ) and touchscreen conditions ( $t(25) = -11.606, p < 0.001$ ). There was not a significant difference in assessment duration between the standardized and touchscreen conditions ( $t(25) = 1.905, p = 0.068$ ). On average the touchscreen assessment condition took the least amount of time to administer.

Table 5. *Descriptive Statistics of Clinical Factors*

Clinical Factor	Experimental Assessment Condition											
	Standardized				Touchscreen				Real-Object			
	Sum	Range	<i>M</i>	<i>SD</i>	Sum	Range	<i>M</i>	<i>SD</i>	Sum	Range	<i>M</i>	<i>SD</i>
Duration	~112	1:23-10:00	4.32	2.18	~91	:59-11:00	3.38	2.14	~343	6:05-23:34	12.69	3.70
PB Trials	56	0-9*	2.07	2.77	37	0-7*	1.37	2.10	48	0-12*	1.78	3.09
NR Trials	41	0-7*	1.52	2.49	8	0-2*	0.30	0.67	14	0-6*	0.52	1.25

Note. Assessment duration is measured in minutes. PB = problem behavior; NR = no-response. \*Range per participant.

### Individual Characteristics

The second aim of this study was to examine how individual characteristics impact performance on assessment conditions. Additional models were tested to examine the effects of subject predictors (i.e., cognition, fine motor skills, maladaptive behaviors, and autism symptomology) on experimental assessment scores. The proportion of reduction in each variance component across models was calculated using pseudo- $R^2$ . Pseudo- $R^2$  represents standardized

effect size and is similar to the traditional  $R^2$ , but accounts for the multiple variance components in crossed random effects models.

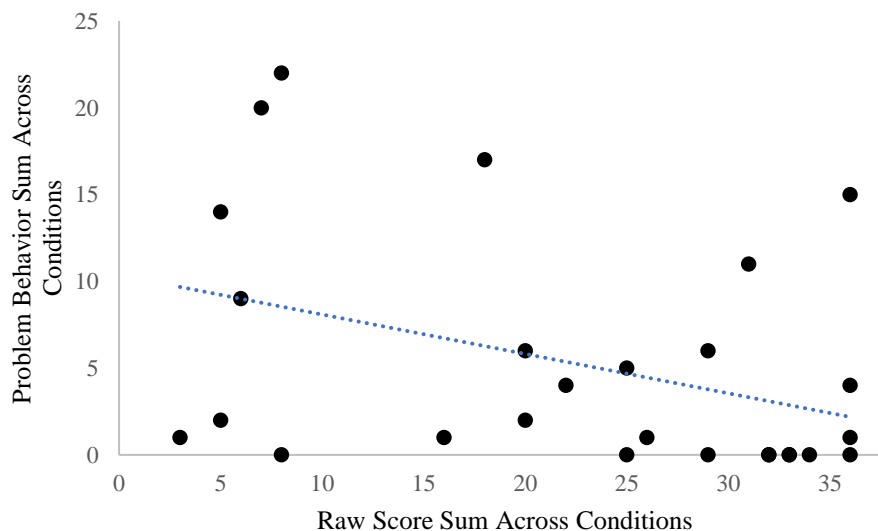
MSEL Visual Reception scores, MSEL Fine Motor scores, and CARS scores were added to the model separately to partition the variance of each individual characteristic. Then scores were added in pairs and groups to calculate the overall proportion of participant variance explained by individual characteristics. Cognition, as measured by the MSEL Visual Reception subtest, reduced participant variance from 5.68 to 3.33. Thus, cognition accounted for 41% of participant variance. Fine motor skills accounted for 37% of participant variance; and autism symptomology, from CARS scores, accounted for 14%. When cognition and fine motor skills were both added to the model, participant variance was reduced by 44%. Altogether, autism symptomology, cognition, and fine motor skills accounted for 44% of participant variance ( $-2\Delta LL(3) = 15, p = .002$ ). Thus, cognition was responsible for 41% of the variance in participants' scores and fine motor skills accounted for 3% of the variance. Autism symptomology did not explain any additional variance in participants' scores that was not already accounted for by cognition and fine motor skills.

Caregiver reports of problem behavior, as measured by the Vineland Maladaptive Behavior subtest, were also added to the model separately. Maladaptive behavior reports increased the variance in participants' scores by 8% when added to the model. This model cannot be compared to the same baseline model as was used to compare the participant characteristics mentioned above because three participants were missing maladaptive behavior scores. A Pearson's correlation coefficient was computed to assess the relationship between maladaptive behavior scores and experimental assessment outcomes for the 24 participants with maladaptive behavior data. There was a very weak negative correlation between caregiver reports of problem

behavior and experimental assessment outcomes, but the correlation was not significant ( $r = -0.24$ ,  $n = 24$ ,  $p = 0.26$ ).

The relationship between the occurrence of problem behavior during the experimental assessment conditions and experimental assessment outcomes was also assessed using a Pearson's correlation coefficient. The overall sum of trials with problem behavior was calculated for each participant by adding the total number of trials with problem behavior in each of the three experimental assessment conditions. Similarly, the overall raw score for each participant was calculated by adding participants' raw scores in each assessment condition. There was a significant, weak negative correlation between occurrence of problem behavior during the experimental assessments and experimental assessment outcomes ( $r = -0.38$ ,  $n = 27$ ,  $p = 0.048$ ). Figure 5 shows a scatterplot of the correlation between problem behavior during experimental assessments and assessment scores.

Figure 5. Occurrence of Problem Behavior in All Experimental Assessment Conditions and Performance Outcomes



## Eye-Gaze

The third aim of this study involved the use of IPLP to examine eye-gaze behavior during the real-object assessment condition. Only incorrect trials were included in this analysis; thus, the five participants who earned a perfect score on the real-object assessment were excluded. Two participants interfered with the cameras during the real-object assessment and their videos could not be coded for eye gaze due to the camera angle. Three trials were excluded from the analysis because the participant's eyes were blocked by the stimuli. If a trial was presented twice (due to no-response or selecting multiple stimuli on the first presentation) only the second presentation was included in the eye gaze analysis.

A total of 92 incorrect trials from 20 participants were examined. A paired-samples t-test was conducted to compare the number of milliseconds spent looking at target stimuli and the average number of milliseconds spent looking at foil stimuli (total number of milliseconds spent looking at foil stimuli divided by 3). There was no significant difference in time spent looking at target stimuli and average time spent looking at foil stimuli during incorrect responses ( $t(125) = -1.473, p = 0.143$ ).

A repeated measures ANOVA was used to examine the number of milliseconds spent looking at target stimuli, average number of milliseconds spent looking at foil stimuli, and number of milliseconds spent looking at non-stimuli during the response period for the 92 trials. There was a statistically significant difference in time spent looking at target stimuli, foil stimuli, and non-stimuli ( $F(2,273) = 13.30, p < 0.001$ ). A post-hoc Bonferroni procedure revealed participants spent significantly longer looking at non-stimuli than at target stimuli ( $t(105) = -3.87, p < 0.001$ ) or foil stimuli ( $t(94) = -3.54, p < 0.001$ ) during incorrect trials.

Of the 92 incorrect trials examined, participants selected one foil stimuli in 73 trials, participants selected more than one stimulus in 8 trials, and in 11 trials the participant did not respond. The 11 no-response trials were examined further using descriptive statistics. The 11 no-response trials came from six different participants. In three of the no-response trials participants did not look at the stimuli at all (participant  $n = 2$ ). In one no-response trial the participant spent longer looking at the target stimuli than at the foils or non-stimuli, and in one trial the participant spent longer looking at foil stimuli. In six of the no-response trials, participants spent more time looking at non-stimuli than at target or foil stimuli (participant  $n = 4$ ). However, when non-stimulus eye-gaze was removed from the analysis for these six trials, there were three trials where participants spent more time looking at the target stimuli than at foil stimuli (participant  $n = 3$ ).

## **Discussion**

It is difficult to gain an accurate measure of receptive language skills in individuals with autism who have limited verbal skills due to the paucity of appropriate assessments for this population. This study examined alternative assessment modalities and procedures for students with autism who had minimal verbal skills by systematically comparing outcomes on three receptive language assessment conditions: a standardized assessment, a touchscreen assessment, and a real-object assessment.

### **Performance on Assessment Conditions**

As hypothesized, participants earned higher scores during the real-object assessment condition than in assessment conditions that used picture stimuli. Participants scored significantly better on the real-object assessment condition than on the standardized assessment

condition, and marginally better on the real-object assessment condition than the touchscreen condition. These findings extend previous research with typically-developing populations showing infants and children perform better on language tasks that use real-objects as opposed to pictures (Cocking & McHale, 1981; Gerhard et al., 2016).

Current literature suggests that real-object stimuli are easier to process and attend to than picture stimuli. Pictures are abstract representations of objects, and the concept of a picture is learned through experience during infancy and childhood (DeLoache et al., 1998). Real-objects are more concrete than pictures, and thus may require less cognitive effort to identify than pictures. Carver and colleagues (2006) found that 18-month-old infants who had typical development were able to identify known stimuli much faster when the stimuli were objects rather than pictures of objects. In addition, real-objects are more salient, more attention-grabbing, and possibly more visually appealing than picture stimuli (Gerhard et al., 2016; Gomez, Skiba, & Snow, 2018). Recent research has shown that people attend to real-objects longer than they attend to pictures of those objects (Gomez, Skiba, & Snow, 2018).

Although the response demand (“Touch X”) was identical across all three conditions, participants’ history and experiences with real-objects, touchscreens, and low-tech picture books may have impacted comprehension of the response demand in each condition. Children are asked to retrieve, pick up, or identify real-objects throughout their day. Similarly, touchscreen devices are becoming ubiquitous in home and school environments. Conversely, recent studies show that children with autism are read to less frequently than their like-aged peers (Westerveld & van Bysterveldt, 2017) and that children with IDD show less interest in picture books than typically-developing peers (Justice et al., 2016). Therefore, participants may have had an easier time



understanding the response demand in the real-object and touchscreen conditions than in the standardized assessment condition.

There were fewer no-response trials during the touchscreen and real-object assessment conditions than during the standardized assessment condition. No-response trials may indicate difficulty comprehending the response demand. It is also possible that some participants had difficulty understanding the response demand in all of the assessment conditions, but due to previous history and experiences, the required response was more inherent with real-objects and touchscreens than with low-tech pictures. Real-objects have affordance, meaning they present an opportunity for physical action, whereas pictures do not afford physical action (Gomez et al., 2018). Participants may have been more inclined to reach out and touch or grab a tangible object that could be physically manipulated than to touch a picture in a book.

Likewise, the interactive nature of the touchscreen device may have impacted participants' responding in this condition. In the touchscreen condition, once a participant selected a stimulus the screen automatically changed to a solid grey screen before the next trial was presented. This was the only assessment condition where the participants' behavior automatically removed the stimuli without any additional assistance from the experimenter. If the task was non-preferred or aversive to participants, the immediate removal of the stimuli after a response may have reinforced participants' responding. In other words, negative reinforcement during the touchscreen condition may have increased responding for some participants.

Similarly, history and previous experiences may have impacted participants' assessment condition preference. Participants' may have associated the objects and touchscreen device with playing games and having fun rather than working; and thus, preferred the real-object and

touchscreen conditions. The low-tech picture booklet may have reminded participants of previous work tasks or testing procedures, making the standardized condition less preferred or even somewhat aversive.

It was also hypothesized that participants would earn better scores on the touchscreen assessment condition than on the standardized assessment condition. Participants performed slightly better in the touchscreen condition than the standardized condition, but the difference was not statistically significant. However, in terms of clinical feasibility, the touchscreen assessment was the easiest to administer and was the most efficient assessment condition.

**Clinical Feasibility.** Overall, the touchscreen assessment condition took the least amount of time to administer and had the fewest occurrences of problem behavior and no-response trials. Conversely, problem behavior and no-response trials occurred most often during the standardized assessment condition. Thus, using high-tech assessment modalities, like touchscreens, may impact the ease and enjoyment of assessments for both the examiner and examinee, but may not make a substantial difference on performance outcomes.

Although there are benefits in using real-objects to assess word-understanding, administration time is a critical factor to consider. On average the real-object assessment condition took about three times longer to administer than the standardized assessment and about four times longer than the touchscreen assessment. The real-object assessment condition contained many items that needed to be manipulated throughout the assessment. After each trial, the experimenter exchanged the objects from the previous trial with four new objects and ensured each object was placed in the correct spot on the board. The number of items and the constant manipulation that was required in the real-object assessment condition greatly increased

administration time. The real-object assessment could have been expedited if the objects in each trial had been glued to smaller individual boards, so the examiner would exchange the board after each trial rather than exchanging each individual object. However, even with individual boards for each trial, the real-object assessment would likely still take longer to administer than the touchscreen and standardized assessments.

It is also important to consider how alternate assessment formats can impact the occurrence and management of problem behavior. The object stimuli in the real-object assessment condition were easy to break, throw, and mouth. Thus, it was more difficult for the experimenter to prevent damage to the many object-stimuli during instances of problem behavior in the real-object condition than it was to prevent damage to the touchscreen device or the picture book during instances of problem behavior in the touchscreen and standardized conditions. Due to the format and duration of the assessment, there were more opportunities to engage in problem behavior in the real-object assessment condition than in the touchscreen or standardized conditions. Yet, the standardized assessment condition had the highest occurrence of problem behavior. These findings show there is promise in using real-objects to assess word understanding in children with autism who have minimal verbal skills; however, more research is needed to make alternative real-object formats efficient and practical for clinical use.

Objective data were not taken on the occurrence of problem behavior by specific topography. However, the experimenter did note that all six topographies that were considered problem behavior for this study (i.e., tantrum, defiance, property destruction, elopement, physical aggression, and self-injury) occurred during the experimental assessments. Subjectively, some problem behavior topographies had a greater impact on assessment administration and outcomes than others. For example, brief instances of defiance that were easily redirected did not result in

the same level of participant or examiner fatigue as brief instances of aggression that were also easily redirected. The intensity and duration of behavior also varied among participants and topographies. For example, some tantrums lasted a few seconds and only consisted of loud whining, but other tantrums lasted longer and included stomping and falling to the ground.

### **Individual Characteristics**

Individual characteristics were also examined to determine how cognition, maladaptive behavior, autism symptomology, and gesture use impacted performance on assessment conditions. Overall, participants had very low cognitive scores, especially when considering the age of the sample. But there was enough variability in cognitive scores to detect main effects that were associated with cognition. Cognition accounted for 41% of the variance in participants' scores. This is not surprising given the link between receptive language skills and cognition (e.g., Mawhood, Howlin, & Rutter, 2000). Although language impairments are sometimes accompanied by cognitive deficits, research has shown that individuals with autism, both with and without verbal skills, have relatively high nonverbal cognitive abilities when compared to receptive and expressive language skills (Joseph, Tager-Flusberg, & Lord, 2002; Plesa Skwerer et al., 2016). Due to the relatively small sample size, interactions between cognition and performance on different assessment conditions were not analyzed. However, future research with a larger sample is needed to determine how individual characteristics such as cognition impact performance on different assessment conditions.

Three percent of the variance in participants' score was unique to fine motor skills. These findings add to previous literature demonstrating a relationship between fine motor skills and receptive language. Both gesture use and nonverbal cognition are strong predictors of language

ability (Luyster et al., 2008). Researchers have also theorized that gross and fine motor skill development facilitates language development (Iverson, 2010). According to this theory, as a child develops more advanced motor skills he or she has more opportunities to interact with people and objects in their environment, which results in more language learning opportunities.

Additionally, it was hypothesized that parental reports of frequent or intense problem behavior would be associated with lower scores across all experimental assessments. However, there was not a significant relationship between parental reports of problem behavior and experimental assessment outcomes. Yet direct observations of problem behavior during the experimental assessments were correlated with lower experimental assessment scores. Peoples' behavior can vary drastically depending on the environment and situation. The majority of the assessments were conducted in schools and the task was highly structured and academic in nature. Thus, it may have been more appropriate to interview teachers about participants' problem behavior rather than parents, because the structure and demands of the assessment were more similar to school environments than home environments. However, data reflecting the occurrence of problem behavior observed during an assessment are likely more closely related to outcomes on that assessment than second-hand reports of problem behavior.

## **Eye Gaze**

The third and final aim of this study was to explore the use of IPLP during incorrect trials in the real-object assessment condition. Results from eye gaze data highlight how prominent attentional difficulties are in this population. During incorrect trials, participants spent more time looking at non-stimuli than at target or foil stimuli. Some examples describing participants' looking behavior during "non-stimuli eye-gaze" include looking at the ceiling, looking at a body

part, such as their hands or feet, looking at the examiner or another person in the room, and looking at the cameras. At times non-stimuli eye-gaze was accompanied by stereotypic behavior such as hand flapping or repeated head shaking.

When no-response trials were examined separately, similar findings of non-stimulus looking emerged. However, in one of the no-response trials the participant looked at the target stimulus longer than the foil stimuli or non-stimuli. This eye-gaze pattern could suggest implicit understanding of the target word. It is also notable that during three of the no-response trials, participants spent the majority of time looking at non-stimuli, but when looking at stimuli, they focused longer on the target stimulus than foil stimuli. This eye-gaze pattern could also suggest that participants knew the correct answer but possibly got distracted and did not respond.

If administration procedures were adjusted to allow more than one repetition of each trial, participants may have made fewer errors due to inattention. In 28% of all repeated trials, participants responded correctly once the trial was repeated. In repeated trials from participants who scored above chance level (raw score of 4 or above) on all assessment conditions, participants responded correctly after the trial was repeated in 52% of repeated trials. In standardized PPVT-4 administration, examiners can repeat trials if the examinee does not respond or appears confused, but the number of repetitions allowed is not specified (Dunn & Dunn, 2007). However, many other standardized measures allow very few repetitions of test trials.

Eye-gaze findings in this study are similar to those from Plesa Skwerer and colleagues (2016) who also found many trials with non-stimulus looking post auditory stimulus. However, Plesa Skwerer and colleagues (2016) used picture-stimuli and eye-tracking technology on a

touchscreen device to measure eye-gaze in slightly older children and adolescents with autism who had minimal verbal skills (Plesa Skwerer et al., 2016). These findings support previous literature about how attentional difficulties effect assessment outcomes in children with autism who have limited verbal skills (Bopp, Mirenda, & Zumbo, 2009; Plesa Skwerer et al., 2016; Tager-Flusberg & Kasari, 2013). It is difficult to get an accurate and reliable measure of word-understanding if examinees do not attend to the stimuli.

### **Clinical Implications**

**Assessment.** Findings from this study support the use of alternative formats to evaluate receptive language skills in children with autism who have minimal verbal skills. Clinicians may gain more robust information about a student's word-understanding when standardized assessment formats and procedures are modified or customized based on the student's skills and interests. Although many direct standardized assessments use picture stimuli to measure word-understanding, when assessing individuals with autism who have limited expressive language, real-objects can be used in replacement of, or in addition to, picture stimuli. However, when using real-objects or modified assessment procedures it is important to allow additional time for the assessment because these modifications increase administration time.

It is also important for clinicians to collect information about a child's word understanding in natural environments. Language is often decontextualized in standardized assessments, meaning language is assessed outside of meaningful, naturally-occurring contexts and activities. Results from standardized measures where language is assessed in a vacuum may offer little information about a child's receptive abilities in everyday situations. In order to create

functional language goals, it is important for clinicians to identify and assess receptive language skills that are needed for the child to succeed in his or her natural environment.

**Intervention.** Many professionals that work with young children with autism use discrete trial training procedures with picture flashcards to teach expressive and receptive word-acquisition. Over the past several years, word-learning apps that function as high-tech flashcards have become increasingly popular (Gosnell, 2011; Hong et al., 2017). Although it is relatively easy for clinicians to take data and find materials for these picture-based discrete trials, it is crucial for therapists to incorporate additional strategies in language-learning intervention. Findings from this study suggest that children with autism may have an easier time identifying and attending to real-objects than pictures of objects. In order to teach new words and concepts, therapists must build on pre-existing knowledge and use materials and situations that are interesting and motivating to the student (Center for Applied Special Technology, 2019). Teaching and modeling new words in natural situations and environments is critical for skill generalization. Also, the use of picture stimuli alone may not be salient enough for some children to learn a new concept or word. Pairing new words or concepts with real-life objects and activities may promote comprehension (Gerhard et al., 2016; Gronneberg & Johnston, 2015).

**Attention.** Due to the attention difficulties seen in many children with autism, it is important to work with an interdisciplinary team to help customize and design instructional tasks and environments to promote attention for students with autism. Occupational therapists can assess the student's sensory needs as well as environmental factors, to ensure the student is getting enough movement and input throughout the day and to eliminate environment distractors like extraneous noise or improper seating (Case-Smith, Weaver, & Fristad, 2015). Nutritional issues such as vitamin deficiency can also impact attention (East et al., 2018). Dieticians can



evaluate a student's nutritional intake and provide strategies to promote a balanced diet.

Physicians and psychiatrists can determine whether pharmacological intervention is necessary to help a student focus during the day or get restful sleep through the night (Earle, 2016; Santosh et al., 2006). Input from special educators and school psychologists is needed to assure content is at the appropriate level given the student's cognitive abilities. Lastly, parents, caregivers, and family members are a crucial component to the interdisciplinary team. These individuals know the student the best and can offer insights on the preferences, dislikes, changes in routine, and other important variables that can impact attention.

## **Limitations**

Although this study offers important information about assessment for individuals with autism who have limited expressive skills, it is important to consider how limitations of this study may impact the outcomes. One limitation of this study is that the characteristics of the participants are not representative of all students with autism who have minimal verbal skills, and thus findings may not generalize to individuals who belong to racial, cultural, and ethnic groups that were not represented in this sample. Specifically, no participants were American Indian and Alaska Native or Native Hawaiian/Pacific Islander. Also, the number of Asian participants in this study is not representative of the general population in the United States. In addition, all but one of the participants were from monolingual English-speaking homes, which does not reflect the cultural and linguistic diversity of many homes in the United States. A larger, more diverse sample would increase the generalizability of this study's findings.

Another limitation is that the stimuli in the real-object condition were not matched for size, color, or material. Objects were chosen based on size, material, color, and similarity of

appearance to standardized pictures. So, for the most part, objects within a trial were similar in size, color, and material; however, some trials contained stimuli of varying sizes. For example, in the “foot” trial, the foil object for ‘neck’ was much larger than the other stimuli in that trial; and in the “mouth” trial, the foil object for ‘eye’ was smaller than the other stimuli in that trial. Some colors and materials may be more visually appealing or attention-grabbing than others. Similarly, looking time can vary based on the size of an object (Libertus et al., 2013). Thus size, color, or material preference could have impacted participants’ responding during the real-object assessment condition. In the future, objects should be created using a 3D printer, so size, color, and material are consistent among stimuli.

In addition, items were presented in a linear array in the real-object condition, as opposed to the quadrant array used in the standardized and touchscreen conditions. The linear stimulus layout may have impacted how participants visually scanned the stimuli in the real-object condition compared to visual scanning in the touchscreen and standardized conditions. Due to the size and fragility of some of the objects, it was not feasible to glue the objects in each trial to individual boards and transport them to and from various assessment locations. A linear array was also more beneficial for eye gaze analysis and camera placement.

Another potential limitation is inconsistent reinforcement across assessment conditions. Participants were given reinforcement in the form of verbal praise, edibles, and preferred items such as sensory toys during or between each assessment condition depending on the child’s needs. The real-object assessment condition required participants to wait much longer between trials than the other assessment conditions. It is possible that the experimenter gave participants more reinforcement to promote waiting between real-object trials than was given in other

conditions. Schedule and amount of reinforcement can impact responding (Kerns & Lanzetta, 1975).

Lastly, due to copyright reasons, the picture-stimuli used in the touchscreen assessment condition were not identical to the pictures in the standardized condition. However, pictures used in the touchscreen condition were matched for size and were chosen to look as similar to the standardized pictures as possible. After data collection began, an updated fifth version of the PPVT was released (PPVT-5; Dunn & Dunn, 2018). The PPVT-5 offers a paper booklet version as well as a digital version that can be conducted on iOS touchscreen devices. Comparison data on performance outcomes between low-tech and high-tech versions is not yet publicly available. However, controlling for picture-stimuli used in the touchscreen and standardized conditions would eliminate confounding variables such as differences in picture clarity or preference among conditions.

## **Future Research**

Many questions remain in terms of receptive language assessment for individuals with autism who have minimal verbal skills. This study shows that the use of real-object stimuli can improve assessment outcomes for some individuals, however, this study only assessed single-word understanding of 11 nouns and one verb. Future research is needed to determine whether real-objects can also be used to assess comprehension of more complex concepts and language structures in connected speech. Many current standardized receptive language measures use pictures to assess concepts such as size and length, as well as comparative concepts, such as order and same/different. Concepts such as semantic relationships between words, and comprehension of connected speech, are often assessed using open-ended and fill-in-the-blank

questions, without any visual stimuli. Examinees need some form of symbolic language to respond to these open-ended and fill-in-the-blank questions. Using objects as opposed to pictures may improve assessment results for some individuals; and pairing open-ended and fill-in-the-blank questions with visual stimuli would allow individuals with minimal verbal skills to respond to more complex receptive language questions.

A study with a larger sample size is needed to determine how individual characteristics, such as cognition, impact performance on different assessment conditions. It was hypothesized that individuals with low cognitive scores will perform better on the real-object assessment condition than on assessments that use pictures. However, a larger sample size is needed to examine interactions between cognition and assessment condition performance. In addition, increasing the number of test items in each condition could increase within-participant variation across assessment conditions, especially for participants with more advanced receptive skills who earn perfect scores on the first 12 test items across conditions. Crossed-random effects models, like those used in the current study, could be used to examine the effects of individual characteristics on assessment condition performance.

There is also a need for exploratory research examining the use of eye gaze during assessment and intervention for individuals with autism who have minimal verbal skills. In this study, eye-gaze data was only collected during the real-object assessment condition. The experimenter traveled to the participants to conduct assessments and portable eye-tracking technology needed for the touchscreen and standardized conditions was not available. Examining eye gaze differences among all three conditions could provide beneficial information on how attention varies based on assessment modality and stimuli. Eye-gaze response latency could be

used to measure attention in each condition. However, it is important to note that data collection and analysis for IPLP procedures used in this study were incredibly time consuming.

High-tech eye-tracking procedures are more efficient and precise than IPLP, and thus should be the focus of future eye-gaze research. Eye-tracking technology uses external devices that sense changes in eye movement from light reflecting off the eye. Algorithms are used to calculate the point of focus on a computer screen or on an image projected to a wall. Wearable eye-tracking devices that provide real-time information on looking patterns have the potential to advance clinical practice. During assessments, clinicians could use eye-tracking technology to make sure the client is focusing on the task before presenting the demand. In addition, during language-learning intervention, eye-tracking technology could help clinicians follow their client's gaze and provide appropriate verbal input based on the client's current focus, rather than relying on clients to follow the clinician's initiations for joint attention. This technology would be especially helpful during sessions in naturalistic settings, like a school playground, where it can be difficult to follow a child's eye-gaze due to the number of visual stimuli present in the environment. However, advances in eye tracking technology are needed before this technology can be used while an individual is moving and in a naturalistic setting.

## References

- Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., ... & Durkin, M. S. (2018). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *MMWR Surveillance Summaries*, 67(6), 1.
- Barbaro, J., & Dissanayake, C. (2012). Developmental Profiles of Infants and Toddlers with Autism Spectrum Disorders Identified Prospectively in a Community-Based Setting. *Journal of Autism and Developmental Disorders*, 42(9), 1939-1948. doi: <http://dx.doi.org/10.1007/s10803-012-1441-z>
- Bavin, E. L., Kidd, E., Prendergast, L., Baker, E., Dissanayake, C., & Prior, M. (2014). Severity of autism is related to children's language processing. *Autism Research*, 7(6), 687-694.
- Bopp, K. D., Mirenda, P., & Zumbo, B. D. (2009). Behavior predictors of language development over 2 years in children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, 52(5), 1106-1120.
- Bouck, E. C., Savage, M., Meyer, N. K., Taber-Doughty, T., & Hunley, M. (2014). High-tech or low-tech? Comparing self-monitoring systems to increase task independence for students with autism. *Focus on Autism and Other Developmental Disabilities*, 29(3), 156-167.
- Brady, N. C., Anderson, C. J., Hahn, L. J., Obermeier, S. M., & Kapa, L. L. (2014). Eye Tracking as a Measure of Receptive Vocabulary in Children with Autism Spectrum Disorders. *Augmentative and Alternative Communication*, 30(2), 147-159. doi: 10.3109/07434618.2014.904923
- Brady, N. C., Thiemann-Bourque, K., Fleming, K., & Matthews, K. (2013). Predicting Language Outcomes for Children Learning Augmentative and Alternative Communication: Child

- and Environmental Factors. *Journal of Speech, Language & Hearing Research*, 56(5), 1595-1612. doi: 10.1044/1092-4388(2013/12-0102)
- Bruckner, C., Yoder, P., Stone, W., & Saylor, M. (2007). Construct Validity of the MCDI-II Receptive Vocabulary Scale Can Be Improved: Differential Item Functioning between Toddlers with Autism Spectrum Disorders and Typically Developing Infants. *Journal of Speech, Language, and Hearing Research*, 50(6), 1631-1638.
- Camilleri, B., & Botting, N. (2013). Beyond static assessment of children's receptive vocabulary: the dynamic assessment of word learning (DAWL). *International Journal of Language & Communication Disorders*, 48(5), 565-581.
- Carver L. J., Meltzoff A. N., Dawson G. (2006). Event-related potential (ERP) indices of infants' recognition of familiar and unfamiliar objects in two and three dimensions. *Developmental Science*. 9 51–62. 10.1111/j.1467-7687.2005.00463.x
- Case-Smith, J., Weaver, L. L., & Fristad, M. A. (2015). A systematic review of sensory processing interventions for children with autism spectrum disorders. *Autism*, 19(2), 133-148.
- Charman, T. (2004). Matching Preschool Children with Autism Spectrum Disorders and Comparison Children for Language Ability: Methodological Challenges. *Journal of Autism and Developmental Disorders*, 34(1), 59-64. doi: 10.1023/B:JADD.0000018075.77941.60
- Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *Journal of Child Language*, 30(1), 213-236. doi: <http://dx.doi.org/10.1017/S0305000902005482>

- Cocking, R. R., & McHale, S. (1981). A Comparative Study of the Use of Pictures and Objects in Assessing Children's Receptive and Productive Language. *Journal of Child Language*, 8(1), 1-13.
- Conti-ramsdén, G., & Durkin, K. (2012). Language Development and Assessment in the Preschool Period. *Neuropsychology Review*, 22(4), 384-401. doi: 10.1111/j.2044-835X.2011.02042.x .
- Dawson, G., & Watling, R. (2000). Interventions to facilitate auditory, visual, and motor integration in autism: A review of the evidence. *Journal of autism and developmental disorders*, 30(5), 415-421.
- DeLoache, J. S., Pierroutsakos, S. L., Uttal, D. H., Rosengren, K. S., & Gottlieb, A. (1998). Grasping the nature of pictures. *Psychological Science*, 9(3), 205-210.  
<http://dx.doi.org/10.1111/1467-9280.00039>
- Dunn, L.M., & Dunn, L.M. (1997). Peabody Picture Vocabulary Test, Third Edition. Circle Pines, MN: American Guidance Service.
- Dunn, L.M., & Dunn, L.M. (2018). Peabody Picture Vocabulary Test, Fifth Edition. Pearson Clinical Assessment.
- Earle, J. (2016). An Introduction to the Psychopharmacology of Children and Adolescents With Autism Spectrum Disorder. *Journal of Child and Adolescent Psychiatric Nursing*, 29(2), 62-71.
- East, P., Delker, E., Lozoff, B., Delva, J., Castillo, M., & Gahagan, S. (2018). Associations Among Infant Iron Deficiency, Childhood Emotion and Attention Regulation, and Adolescent Problem Behaviors. *Child Development*, 89(2), 593-608.



- Edwards, S., Fletcher, P., Garman, M., Hughes, A., Letts, C. & Sinka, I. (1997). Reynell Developmental Language Scale – III. Windsor: NFER-Nelson.
- Ellis Weismer, S., Lord, C., & Esler, A. (2010). Early Language Patterns of Toddlers on the Autism Spectrum Compared to Toddlers with Developmental Delay. *Journal of Autism and Developmental Disorders*, 40(10), 1259-1273. doi: <http://dx.doi.org/10.1007/s10803-010-0983-1>
- Farmer, C., Golden, C., & Thurm, A. (2016). Concurrent validity of the differential ability scales, with the Mullen Scales of Early Learning in young children with and without neurodevelopmental disorders. *Child Neuropsychology*, 22(5), 556-569.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., ... Reilly, J. S. (1993). The MacArthur Communicative Development Inventories: User's guide and technical manual. San Diego, CA: Singular.
- Gerhard, T. M., Culham, J. C., & Schwarzer, G. (2016). Distinct Visual Processing of Real Objects and Pictures of Those Objects in 7- to 9-month-old Infants. *Frontiers in Psychology*, 7, 827. <http://doi.org/10.3389/fpsyg.2016.00827>
- Golinkoff, R., Hirsch-Pasek, K., Cauley, K., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, 14, 23-46.
- Goodwin, A., Fein, D., & Naigles, L. R. (2012). Comprehension of wh-questions precedes their production in typical development and autism spectrum disorders. *Autism Research*, 5(2), 109-123.

- Gomez, M. A., Skiba, R. M., & Snow, J. C. (2018). Graspable Objects Grab Attention More Than Images Do. *Psychological Science*, 29(2), 206–218.  
<https://doi.org/10.1177/0956797617730599>
- Gosnell, J. (2011). Apps: An Emerging Tool for SLPs: A plethora of apps can be used to develop expressive, receptive, and other language skills. *The ASHA Leader*, 16(12), 10-13.
- Gronneberg, J., & Johnston, S. (2015, April 6). 7 things you should know about universal design for learning [Brief]. Educause Learning Initiative. Retrieved from <http://www.educause.edu/library/resources/7-things-you-should-know-about-universal-design-learning>.
- Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. New York, NY: Routledge.
- Hong, E. R., Gong, L. Y., Ninci, J., Morin, K., Davis, J. L., Kawaminami, S., ... & Noro, F. (2017). A meta-analysis of single-case research on the use of tablet-mediated interventions for persons with ASD. *Research in developmental disabilities*, 70, 198-214.
- Hudry, K., Leadbitter, K., Slonims, V., McConachie, H., Aldred, C., Howlin, P., . . . Consortium, t. P. (2010). Preschoolers with autism show greater impairment in receptive compared with expressive language abilities. *International Journal of Language & Communication Disorders*, 45(6), 681-690. doi: 10.3109/13682820903461493
- Iverson J. M. (2010). Developing language in a developing body: the relationship between motor development and language development. *Journal of Child Language*, 37(2), 229–261.  
doi:10.1017/S0305000909990432

- Joseph, R. M., Tager-Flusberg, H., & Lord, C. (2002). Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. *Journal of child psychology and psychiatry, and allied disciplines*, 43(6), 807–821.
- Justice, L. M., Logan, J. A., Işitan, S., & Saçkes, M. (2016). The home-literacy environment of young children with disabilities. *Early Childhood Research Quarterly*, 37, 131-139.
- Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research*, 6(6), 479-493. doi: <http://dx.doi.org/10.1002/aur.1334>
- Kerns, C., & Lanzetta, John T. (1975). Effects of schedule and amount of observed reinforcement on response persistence. *Journal of Personality and Social Psychology*, 31(6), 983-991.
- Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An Investigation of Language Impairment in Autism: Implications for Genetic Subgroups. *Language and Cognitive Processes*, 16(2-3), 287-308.
- Libertus, K., Gibson, J., Hidayatallah, N. Z., Hirtle, J., Adcock, R. A., & Needham, A. (2013). Size matters: How age and reaching experiences shape infants' preferences for different sized objects. *Infant Behavior and Development*, 36(2), 189-198.
- Luyster, R. J., Kadlec, M. B., Carter, A., & Tager-flusberg, H. (2008). Language Assessment and Development in Toddlers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 38(8), 1426-1438. doi: <http://dx.doi.org/10.1007/s10803-007-0510-1>
- Mawhood, L., Howlin, P., & Rutter, M. (2000). Autism and developmental receptive language disorder—A comparative follow-up in early adult life. I: Cognitive and language

- outcomes. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 41(5), 547-559.
- McConachie, H., Parr, J. R., Glod, M., Hanratty, J., Livingstone, N., Oono, I. P., ... & Garland, D. (2015). Systematic review of tools to measure outcomes for young children with autism spectrum disorder. *Health Technology Assessment*, 19(41).
- Mullen, E. (1995). Mullen Scales of Early Learning: AGS Edition. Circle Pines, MN: AGS.
- Muller, K., & Brady, N. (2016). Assessing Early Receptive Language Skills in Children With ASD. *Perspectives of the ASHA Special Interest Groups*, 1(1), 12-19. doi: 10.1044/persp1.SIG1.12
- Pickard, K. E., & Ingersoll, B. R. (2015). Brief Report: High and Low Level Initiations of Joint Attention, and Response to Joint Attention: Differential Relationships with Language and Imitation. *Journal of Autism and Developmental Disorders*, 45(1), 262-268.
- Plesa Skwerer, D., Jordan, S., Brukilacchio, B., & Tager-flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism : the international journal of research and practice*, 20(5), 591-604. doi: 10.1177/1362361315600146
- Preissler, M. A. (2008). Associative learning of pictures and words by low-functioning children with autism. *Autism*, 12(3), 231-248.
- Prelock, P.A. & McCauley, R.J. (2012). *Treatment of Autism Spectrum Disorders*. Baltimore, MD: Paul H. Brookes Publishing Co.

- Robinson CW, Shore WJ, Hull Smith P, Martinelli L. Developmental differences in language comprehension: What 22-month-olds know when their parents are not sure. Poster presented at the International Conference on Infant studies; Brighton. July, 2000.
- Santosh, P., Baird, G., Pityaratstian, N., Tavaré, E., & Gringras, P. (2006). Impact of comorbid autism spectrum disorders on stimulant response in children with attention deficit hyperactivity disorder: A retrospective and prospective effectiveness study. *Child: Care, Health and Development*, 32(5), 575-583.
- Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). Childhood autism rating scale, (CARS2). Torrance, CA: WPS.
- Sparrow, S., Cicchetti, D., & Balla, D. (2005). *Vineland Adaptive Behavior Scales*, Second Edition. Livonia, MN: Pearson.
- Swensen, L. D., Kelley, E., Fein, D., & Naigles, L. R. (2007). Processes of language acquisition in children with autism: Evidence from preferential looking. *Child development*, 78(2), 542-557.
- Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research*, 6(6), 468-478.
- Tager-flusberg, H., Plesa Skwerer, D., Joseph, R. M., Brukilacchio, B., Decker, J., Eggleston, B., . . . Yoder, A. (2016). Conducting research with minimally verbal participants with autism spectrum disorder. *Autism: the international journal of research and practice*, 1-10.
- Tomasello, M., & Mervis, C. B. (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development*, 59(5), 174-179. doi: 10.1111/j.1540-5834.1994.tb00186.x

- Vaughn, B. J., & Horner, R. H. (1997). Identifying instructional tasks that occasion problem behaviors and assessing the effects of student versus teacher choice among these tasks. *Journal of Applied Behavior Analysis*, 30(2), 299-312.
- Venker, C. E., Eernisse, E. R., Saffran, J. R., & Weismer, S. E. (2013). Individual differences in the real-time comprehension of children with ASD. *Autism Research*, 6(5), 417-432. doi: <http://dx.doi.org/10.1002/aur.1304>
- Volden, J., Smith, I. M., Szatmari, P., Bryson, S., Fombonne, E., Mirenda, P., . . . Thompson, A. (2011). Using the Preschool Language Scale, Fourth Edition to Characterize Language in Preschoolers With Autism Spectrum Disorders. *American Journal of Speech-Language Pathology*, 20(3), 200-208.
- Westerveld, M., & van Bysterveldt, A. K. (2017). The Home Literacy Environment of Preschool-Age Children with Autism or Down Syndrome. *Folia Phoniatrica et Logopaedica*, 69(1-2), 43-53.
- Wiig, E.H., Secord, W., & Semel, E. (1992). Clinical Evaluation of Language Fundamentals - Preschool San Antonio: The Psychological Corporation, Harcourt Brace & Co.
- Wiig, E. H., Semel, E., & Secord, W. A. (2013). Clinical Evaluation of Language Fundamentals—Fifth Edition (CELF-5). Bloomington, MN: NCS Pearson.
- Zimmerman, I., Steiner, V., & Pond, R. (2002). Preschool Language Scale, Fourth Edition. San Antonio, TX: The Psychological Corporation.
- Zimmerman, I. L., Steiner, V. G., & Pond, E. (2011). Preschool Language Scales- Fifth Edition (PLS-5). San Antonio, TX: Pearson.

## Appendix A

### Caregiver Questionnaire

#### *Gesture Use*

1. Will your child push an object or your hand away to show you they don't want something?
  - a. Never
  - b. Sometimes
  - c. Often
2. Does your child reach toward you or toward objects that they want?
  - a. Never
  - b. Sometimes
  - c. Often
3. Does your child guide your hand or pull you towards a desired item?
  - a. Never
  - b. Sometimes
  - c. Often
4. Will your child point to an object that is within reach to show you they want something?
  - a. Never
  - b. Sometimes
  - c. Often
5. Will your child point to an object that is out of reach to show you they want something?
  - a. Never
  - b. Sometimes
  - c. Often
6. Will your child give you an item that is broken or that they need help opening?
  - a. Never
  - b. Sometimes
  - c. Often
7. Does your child wave to say hello or goodbye?
  - a. Never
  - b. Sometimes
  - c. Often
8. Does your child nod or shake their head to indicate yes and no?
  - a. Never
  - b. Sometimes
  - c. Often

#### *Touchscreen History*

9. How often would you say your child interacts with touchscreen devices, like iPads, tablets, and phones?
  - a. Never
  - b. Rarely
  - c. Sometimes
  - d. Often
10. Why does your child interact with these devices: (select all that apply)

- a. To play games, watch videos, or listen to music
  - b. To complete work or school tasks (e.g., learning apps, visual schedules, self-monitoring)
  - c. To take pictures
  - d. To communicate (e.g., AAC device)
  - e. Other:
11. What do they primarily do on touchscreen devices? (select one)
- a. Play games, watch videos, or listen to music
  - b. Complete work or school tasks (e.g., learning apps, visual schedules, self-monitoring)
  - c. Take pictures
  - d. Communicate (e.g., AAC device)
  - e. Other:
12. On a scale of 1 to 4, with 1 being not preferred and 4 being highly preferred, how would you rate your child's preference for touchscreen devices?
- a. 1 (not preferred)
  - b. 2 (somewhat preferred - the child has many other items/activities that are more preferred than touchscreen devices)
  - c. 3 (moderately preferred - the child has a few items/activities that are more preferred than touchscreen devices)
  - d. 4 (highly preferred - touchscreens are the child's most preferred item/activity)



## Appendix B

### Screening Questions

1. What is your child's date of birth?
2. What is your child's primary diagnosis and do they have any coexisting diagnoses?
3. How many words can your child say (verbal, sign, SGD, PECS)?
4. Is your child's hearing within or corrected to be within normal limits?
5. Is your child's vision within or corrected to be within normal limits?

#### *Attention*

6. How long will your child attend to a non-preferred, work-like task?
7. What are some things you (or the child's teacher) do to help them pay attention for longer periods of time?
8. How long will your child attend to a preferred task?

#### *\*Problem Behavior*

9. Does your child engage in problem behavior, for example self-injury, aggression, running away etc.? (if no, skip to #15)
10. What does the behavior look like (ask for each behavior mentioned)?
11. How often do they engage in these behaviors (ask for each behavior mentioned)?
  - a. Rarely
  - b. Sometimes
  - c. Often
12. When do they engage in these behaviors?
13. Is there anything you (or the school) do to help reduce or prevent the occurrence of problem behavior?

#### *\*Preferred Items/Activities*

14. For this study, your child will be given three short assessments. He/she will be given breaks between each assessment what are some things your child likes to do during their breaks (e.g., play on iPad, tickles, play basketball, play with a bumble ball, stim, listen to music, sit in silence, etc.)?
15. What are things your child likes to work for (e.g., a break, iPad, stickers, food, trampoline, tickles, etc.)?
16. Does your child have any dietary restrictions?
17. If so, what are their dietary restrictions?

\* Questions from these sections will be administered to parents/caregivers as well as the classroom teacher, if the assessments are done in the school.

## Appendix C

### Data Sheet

**Participant ID:** \_\_\_\_\_

**Date:** \_\_\_\_\_

**Assessment:** Standardized   Real-Object   Touch-Screen

**Start Time:** \_\_\_\_\_

**End Time:** \_\_\_\_\_

Training					RD	PB	NR
1	Puppy	Chair	Boy	Bike			
2	Puppy	Chair	Boy	Bike			
3	Puppy	Chair	Boy	Bike			
4	Puppy	Chair	Boy	Bike			

Test Items					RD	PB	NR
1	Rose	Pumpkin	Ball	Bird			
2	Dog	Donut	Comb	Bed			
3	Spoon	Fish	Cake	Cup			
4	Shoulder	Hand	Foot	Neck			
5	Calf	Sheep	Duck	Horse			
6	Lamb	Wheel	Bib	Banana			
7	Sandwich	Shoe	Bear	Stool			
8	Umbrella	Bee	Pretzel	Cup			
9	Chess	Eating	Swinging	Knocking			
10	Elephant	Kite	Broom	Bus			
11	Cube	Pizza	Starfish	Flower			
12	Mouth	Eye	Ear	Nose			

## Appendix D

### Operational Definitions for Problem Behavior

Aggression: The child is physically aggressive to another individual. The child hits, kicks, pushes, bites, grabs, pulls, scratches, shoves, squeezes, spits at, pinches, or throws an object at another individual.

Elopement: The child runs or walks away from their chair or the testing room.

Self-Injury: The child engages in behaviors that could cause injury to themselves; the child is aggressive to themselves: biting self, hitting self OR hitting self against something (e.g. wall or object), scratching or picking at body part, poking own eye, pulling own hair, pica (consuming an inedible substance such as dirt), and mouthing (child places toy, clothes, or other item in mouth not for the purpose of learning or exploration).

Property Destruction: The child throws, breaks, or destroys their own property or others' property.

Defiance: The child excessively says "no" or refuses to obey authority, or refuses to answer a question or participate in the assessment.

Tantrum: The child has an emotional outburst or meltdown, may involve crying or falling to the floor.

**\*Note:** Behavior mentioned above will be marked as PB in the green column next to the appropriate trial. Any attempts of these behaviors will also be marked as PB in the same manner. Examples of attempts: If the participant tries to run out of the testing area but the examiner blocks the exit, this counts as attempted elopement and will be marked as PB. If the participant tries to hit their head and the examiner or paraprofessional blocks the child's hand from making contact with their head, this counts as attempted self-injury and will be marked as PB.

## Appendix E

### Fidelity Checklist

Participant ID:

Date:

Your Initials:

Assessment: standardized, touch-screen, real object (circle one)

<b>Target</b>	<b>Presented in correct order?</b>	<b>Experimenter presented verbal label when participant is looking at her or the stimuli?</b>	<b>Presented the verbal label only 1 or 2 times?</b>	<b>Refrained from prompting – including gesture, body positioning, eye gaze?</b>
<b>Ball</b>	Yes No	Yes No	Yes No	Yes No
<b>Dog</b>	Yes No	Yes No	Yes No	Yes No
<b>Spoon</b>	Yes No	Yes No	Yes No	Yes No
<b>Foot</b>	Yes No	Yes No	Yes No	Yes No
<b>Duck</b>	Yes No	Yes No	Yes No	Yes No
<b>Banana</b>	Yes No	Yes No	Yes No	Yes No
<b>Shoe</b>	Yes No	Yes No	Yes No	Yes No
<b>Cup</b>	Yes No	Yes No	Yes No	Yes No
<b>Eating</b>	Yes No	Yes No	Yes No	Yes No
<b>Bus</b>	Yes No	Yes No	Yes No	Yes No
<b>Flower</b>	Yes No	Yes No	Yes No	Yes No
<b>Mouth</b>	Yes No	Yes No	Yes No	Yes No